A Study of the Eysenck-Furneaux Approach to the Analysis of Performance on Intelligence Tests.

Berger, M

A Study of the Eysenck-Furneaux Approach to the Analysis of Performance on Intelligence Tests.

Michael Berger

A thesis submitted for the degree of

Doctor of Philosophy in Psychology.

Institute of Psychiatry, University of London.

January 1976.

PAGE NUMBERS CUT OFF IN ORIGINAL

ABSTRACT.

The studies reported here derive from the attempts of Eysenck and Furneaux to overcome some of the many limitations which inhere in the construction, application and interpretation of intelligence tests.

In the first section of this thesis, some of the specific deficiencies of intelligence tests are identified and discussed. This is followed by a critical examination of the concept and measurement of speed. It is concluded that both are inadequate and that despite much research, we are not yet able to appreciate the role of speed in test performance.

Studies of reaction time are examined in relation to both intelligence and personality, and detailed consideration is given to the viability of an information theory interpretation of reaction time. Problems of methodology and interpretation are identified and these are found to militate against unequivocal conclusions.

Finally, the concepts of accuracy and continuance and associated empirical studies are considered. Again, unequivocal conclusions fail to assert themselves.

The second section describes Furneaux's analysis of problem solving and the attempt to utilise his difficulty scaling procedures on data gathered using computer administered tests (specially arranged versions of the Mill Hill and Matrices). In the course of presenting his approach, certain practical and empirical limitations are identified. It was found that only limited scaling was possible but that for a proportion of the items of both tests, there was a strong linear relationship between time difficulty and item solution speed.

The present study failed to find evidence of a relationship between reaction time data transformed to an information theory based index and mental speed. Only equivocal evidence was obtained when these indices were correlated with measures of intelligence. Finally, some of the findings were consistent with predictions derived from Eysenck's theory of personality.

To my parents,
to Zillah and Jacob, and
especially to Sarah.

ACKNOWLEDGEMENTS.

I wish to express my sincere thanks to my supervisor, Prof. H. J. Eysenck, who not only suggested that I persue the research reported here, but also for enabling me to use the full facilities of the Department of Psychology at the Institute of Psychiatry. I appreciate as well his patience and gentle persistence in encouraging me to complete this work. My thanks are also due to Dr. S. B. G. Eysenck for her encouragement.

To Owen White I owe a special gratitude. Without his advice on the statistical analysis of the data, I doubt that the study would have been completed. His advice, his patience, and his friendship over many years are very sincerely appreciated.

A number of people in the Psychology Department have assisted me in various ways, and I wish to acknowledge my indebtedness to them. Mr. F. Brown, Mr. R. Bluffield, and Mr. L. Law constructed the apparatus used in this study and engineered its attachment to the computer; Mrs. V. Beevers always helped with card punching and Miss. N. Hemsley with some of the very many computations; Mr. F. Lillie assisted in testing the people who acted as subjects.

CONTENTS.

LIST OF FIGURES.

LIST OF FIGURES (Contnd.).

LIST OF TABLES.

LIST OF TABLES (Contnd.).

LIST OF TABLES (Contnd.).

APPENDIX.

SECTION A.

This thesis is presented in two major sections. The first is concerned with setting a theoretical and empirical framework for the research reported in Section B.
This section provides an introduction to the research problems and then considers some important problems in contemporary intelligence measurement, as well as some of the solutions proposed by various research workers. Special emphasis is given to the measurement of difficulty and to the concept and measurement of speed. The relationship between speed and intelligence, and between speed and personality is then discussed. The section is concluded with an examination of the literature on persistence and accuracy.

I. INTRODUCTION: A STATEMENT OF THE RESEARCH PROBLEMS.

In a paper published in 1967 Eysenck (1967a) presented a number of major criticisms of contemporary approaches to the study of intelligence. This paper brought together the theoretical arguments and empirical evidence for his view that psychometrics had "become almost completely divorced" from the mainstream of experimental psychology. Research had become overly dependent on factor analysis which could not on its own provide the answers that psychologists sought in their study of intelligence. Eysenck further asserted that the basis of intelligence measurement, the total test score, confounded a number of components and was thus unsuitable as the basic unit for the analysis of test performance. Finally, he noted the lack of concern with "mental speed" and suggested that it should be "restored to its theoretical pre-eminence as the main cognitive determinant of mental test scoring ability".

The psychological study of intelligence has many facets, ranging from problems of definition, measurement and structure to such issues as inheritance and the social implications of intelligence testing. The decades which followed the establishment of intelligence as an important psychological concept have, with few exceptions, witnessed major and minor controversies about one or other of its facets. While it is unlikely that many of the problems will be answered in the foreseeable future, some of the more fundamental issues have at least been identified. Butcher (1968) has pointed out that from the point of view of the scientific study of intelligence, what is needed is a law or set of laws which can act as the basis for major advances in our understanding of intelligence. Such laws would, according to Butcher, help establish an acceptable definition of intelligence which would then facilitate further developments. The work of Furneaux (1961) was seen as a potentially important approach in relation to this problem, although at the time he made his comments Butcher was not able to assess the significance of Furneaux's work. As he says -

> "It is not immediately obvious whether or not this finding (of Furneaux's)* is of far-reaching importance for the purposes we have been considering, nor even (without replication in other experiments) whether it attains the status of a psychological law at all, but it certainly cries out for fuller investigation". (Butcher,1968, p.35-36).

*My addition

After presenting his criticisms of contemporary approaches to the study of intelligence, Eysenck (1967a) outlined the main features of his own approach. Of central importance to it is Furneaux's (1961) work. Furneaux's difficulty scaling procedures not only take account of some of the criticisms made of the total score unit, but also show that difficulty and log time to correct solution are linearly related. The import of this relationship is expressed as follows -

> "The increase of log latency with increase in item difficulty turns out to have the same slope for all individuals tested, and is thus a <u>constant</u>, one of the few which exist in psychology".
> (Eysenck 1967a).

Despite the obvious significance of this finding, there were no attempts at replication until the late 1960's, even though Furneaux had begun to publicise his work as early as 1948 (Furneaux 1948,1950,1952). This delay was partly due to the length of time it took for the actual procedures to be published (Furneaux 1961) and because, as Butcher (1968) remarked, the paper which reported this main finding was "decidedly obscure" in its presentation.

Eysenck's approach to the measurement of intelligence is, as he recently noted (Eysenck 1973a p.x), a departure from traditional psychometrics. As such, its viability is dependent on the support it receives from adequate validation studies. In particular, the theoretical and empirical superstructure needs to be well supported at its base. It is for this reason that the Furneaux model and the empirical relationships which it generates need to be critically evaluated, especially by empirical studies aimed at testing its resillience under different conditions. A major purpose of the present investigation is to provide such an evaluation. Although there have been studies supporting some elements of the Eysenck-Furneaux approach (e.g Farley 1966) only two studies known to the present writer have attempted a direct test of the Furneaux model. The present study if one of these. The second was carried out by Brierley (1969). Both studies were run concurrently (see Brierley 1969, p.164) although the latter included tests of the Furneaux model as part of a broader investigation otherwise unrelated to the present study.

An associated feature of Eysenck's (1967a) approach is the attempt to reinstate 'speed' as a major variable in the study of intelligence. The relationship between log. latency and item difficulty derived by Furneaux serves to buttress the importance of speed but further evidence is needed, together with a theoretical framework which can serve to link problem-solving speed with other measures of quickness. A theoretical link was suggested by Furneaux in the same paper which presented the problem-solving model.

Furneaux treats problem-solving (item solution) as a special case of multiple-choice reaction time. The problem solving process requires, as one of its components, a search for the solution, As conceptualized by Furneaux, the search process involves a series of operations each component of which occupies a fixed time. The more difficult the problem, the greater will be the number of components required and thus an increase in the time required for the search. Furneaux developed this model on the basis of work described by Hick (1952), Hyman (1953) and others. These studies had demonstrated that choice reaction time was linearly related to the complexity of the choice situation. This complexity can be transformed into an information-theory unit, the bit. In the case of simple reaction time, 0 bits of information are involved. One bit of information is present when two alternatives are presented, two bits when there are four possibilities, three bits when there are eight choices, and so on. By using a choice reaction time task involving 0 to 2 or more bits, it is possible to compute a 'rate of gain' measure which is the slope of the best fit line through the mean reaction time for each set of choices.

While simple reaction time does not show a strong relationship with intelligence, there is evidence of a relationship between intelligence and the rate of gain measure in choice reaction time (Roth 1964). This evidence is seen by Eysenck (1967a) as indicating that reaction time experiments do not, as previous studies suggested, contradict a theory in which speed is a central concept. Also, Roth's research has important implications for the choice reaction time model of problem solving put forward by Furneaux. At the time that the present study was instigated, Roth's (1964) research had not been repeated; nor had there been any attempt to examine the relationship between 'speed' as defined by Furneaux and a measure of rate of information processing. The investigation of such relationships is however crucial if problem solving speed is to be given the status accorded it by Eysenck. Hence, a further problem for investigation in the present study is the nature and strength of the relationship between problem solving speed and 'rate of gain' of information.

It is generally recognised that non-cognitive factors can influence performance on tests of intelligence (Gaudry and Spielberger 1971;Naylor 1972). However, there are few major theories of personality which make explicit the role of personality in test performance. Similarly, there are few substantial theories of intelligence which try to incorporate personality variables in their formulation. Two outstanding exceptions however are Eysenck (1967a) and Cattell (1971). Both have consistently attempted to integrate these two areas of psychology although Cattell's approach does not as readily generate testable predictions about test performance as does Eysenck's (see Eysenck 1967a, for example). One of the important features of Furneaux's model is it's inclusion of personality parameters in it's formal structure.

Furneaux (1956) has published a number of tests based on his theoretical ideas and these tests have been applied to different personality groups to test hypotheses about personality-performance relationships (Farley 1966a+b). However, there has never been a detailed examination of the relationship between personality variables and the parameters derived from a direct application of the Furneaux model. Furneaux did not use his difficulty scaling procedures to scale the items of his published tests (Brierley 1969). Thus, while both Furneaux (1961) and Eysenck (1967a,1973a) explicitly recognise the importance of personality factors in test performance there is a need to investigate such relationships directly in relation to the measures derived in the course of difficulty scaling. This was the third main goal of the present study.

The research reported in this thesis was conceived and developed in 1967 and completed by mid-1968. It was carried out in the Department of Psychology at the University of London Institute of Psychiatry as part of the department's ongoing programme of investigations into the measurement of intelligence.

II. THE CONTEXT

I. The Form of Intelligence Tests.

Mental testing, as a branch of psychology, grew out of experimental psychology. There was, as Boring (1957) notes " ..... no schism in the 1890's" (p.571). Yet within the next two decades the mainstream fragmented to the extent that experimental psychology and mental testing now require separate historical analyses. As Boring (1957) puts it: "And does the history of mental testing belong in the history of experimental psychology? Not really ....." (p.570). The events that led to the separation constituted what Kuhn (1971) might regard as a minor revolution in the history of psychology as a science, even though the revolution itself was technological rather than scientific. Taking Kuhn's (1971) analysis further, it can be suggested that at the turn of the century, mental testing was in a state of crisis, at least in so far as the psychology of intelligence was concerned. This aura of crisis is well illustrated by Spearman's (1904b) review of the early attempts to relate mental test scores to various criteria of academic competence.

> "Thus far, it must be confessed, the outlook is anything but cheerful for experimental psychology in general. There is scarcely one positive conclusion concerning the correlation between mental tests and independent practical estimates that has not been with equal force flatly contradicted; and amid this discordance, there is a continually waxing inclination - especially among the most capable workers and exact results - absolutely to deny any such correlation at all."
>
> (Spearman 1904b)

The early crisis was helped towards a resolution by a number of important papers. Two were published by Spearman (1904a,b) and one by Binet and Simon in 1905. Together, these papers had a number of ramifications, perhaps the most important of which were the foundations that they laid for applied intelligence testing, psychometrics, and the various theories of intelligence that have been associated with correlational psychology.

With certain exceptions, notably Binet and Ebbinghaus, most of the tests used in the early period of mental testing were single tests directed mainly at measuring sensory and motor processes and memory. These tests were derived from experimental psychology laboratories and were used in the study of inheritance (exemplified in the work of Galton), in the prediction of academic success or in other aspects of research on individual differences (Boring 1957; Freeman 1939).

The tests used by Ebbinghaus and those being developed by Binet were aimed at the 'higher mental processes'. By 1898, according to Boring (1957), it was possible to show that insofar as item content was concerned, the Binet-type item had "won out". However, not everyone was prepared to concede a victory to Binet. Spearman (1904b) after reviewing the work of Binet and Henri published in 1895, noted an unacceptable "new feature" in their tests:

> "Hitherto, these had been of the most elementary and unequivocal nature possible, as befits the rigour of scientific work.....Binet and Henri appear now to seek tests of a more intermediate character, sacrificing much of the elementariness, but gaining greatly in approximation to the events of ordinary life. The result would seem likely to have more practical than theoretical value".

In this early period, the tests were not organised into scales, and if a number of them were administered concurrently, the scores were not combined (Freeman 1939). The development of a scale which produced a total score is generally attributed to Binet and Simon. However, as Wolf (1973) indicates in her biography of Alfred Binet, the idea probably originated in the set of tests used by Blin and Damaye for their diagnosis of subnormality. Their tests, produced before the 1905 scale, yielded a total score and were sufficiently developed to provide a crude range indicating 'normal' performance.

The development of the first scale of intelligence was not a "fortuitous event" (Wolf 1973). For over two decades, Binet had been searching for a procedure which would enable a discrimination to be made between the then accepted medical classification of three grades of mental deficiency. The central idea of the 1905 scale published by Binet and Simon was that individuals of different degrees of intelligence at the same level of maturity could be distinguished by the number of tests they can pass. The component tests of this scale were graded according to difficulty level and the notion of an age-scale was already present (Freeman 1939).

Taken separately, neither the content of the items nor the idea of a scale was novel. The novelty was to be found in their combination and in the provision of explicit instructions for administration and scoring. "What had been a 'test' now became a sub-test or item of a scale which as a totality yielded a composite measurement of a complex function". (Du Bois 1970,p.36).

On the 1908 scale and later editions of the Binet,mental age was computed by first finding the basal age and then adding one year for each five items passed above that level. This procedure of adding together disparate items to produce an agglomerate score is of paramount importance because it represents an apprcach to quantification which has affected the scoring of psychological tests ever since.

Further early developments served to consolidate some of the characteristics of early tests. The first point scale was developed by Yerkes,Bridges and Hardwick. Instead of mixing items at one age level, it grouped together items of homogeneous content graded in difficulty within sub-tests. The total score was then related to a table of norms from which a mental age was obtained. In principle, it did not differ from the Binet tests.

A second important development, group tests, also occurred in this early period. Although Otis is credited with this 'invention', there is some controversy as to who was responsible (Linden and Linden 1968). As an innovation, group testing facilitated data collection, an advantage that was fully exploited during the First World War (Boring 1957; Freeman 1939; Du Bois 1970). As will be seen later, group tests also created certain problems which have tended to be ignored in investigations of intelligence.

2. Some Inadequacies of Intelligence Tests.

The inadequacies of intelligence tests have been commented on periodically. Thorndike et al.(1927), while recognising that the then current tests were an improvement on their predecessors, nevertheless recognised three sources of deficiency. According to their view, the tests were ambiguous in content, their units arbitrary and the results of uncertain significance. In 1925 Thorndike had suggested that ability should be analysed into level, range and speed. It was acknowledged that for practical purposes, a test battery which combined speed,level and range (or extent) in unknown amounts might well be useful. However,

> "For rigorous measurements.....it seems desirable to treat these three factors separately, and to know the exact amount of weight given to each when we combine them". (Thorndike et al.1927, p.25).

Furneaux's (1961) ancestry is readily discernable in this statement, although the components he selected as important differed somewhat from those emphasised by Thorndike. Like Thorndike, Furneaux also recognised the practical utility of the combination of 'unknown amounts' of the components.

For Thorndike, the measurement of extent and speed posed no problems. The former could be accomplished by simply counting the number of tasks correct in a sample of problems from all areas of intellectual functioning. Speed too is readily measured. The measurement of difficulty however, posed the greatest problem. If difficulty could be properly measured,'level' would readily succumb to measurement, and it would then also be possible to quantify 'extent' and 'speed' at any level. Hence, for Thorndike, the quantification of difficulty represented the central task in the measurement of intelligence. This problem will be considered later.

In 1927 Spearman criticised the "hotchpotch" procedure employed in scoring intelligence tests. What he failed to see was the relevance of his comments for the tests of homogeneous content which he used. Spearman's concern was the crudity of the procedure inherited from Binet, "the prevalent procedure of throwing a miscellaneous collection of tests indiscriminately into a single pool..."(p.71). Spearman, like Thorndike, also recognised the need, as part of a scientific approach, to "dissect" the subject matter under investigation. The problems entailed in the use of an agglomerate score, and the need for 'dissection' have been re-emphasised in the more recent critical analyses of intelligence tests put forward by Furneaux (1961) and Eysenck (1967a,1973a).

The fundamental criticism advanced by Eysenck concerns the assumption that equal scores on a test are intellectually equivalent. According to common usage, it is assumed that if individuals obtain the same total score on a test, then such scores have equivalent psychological significance. While this may be true in a probabilistic sense, if the scores were arrived at in different ways, such equivalence is questionable.

Eysenck (1967a) illustrates the problems in the following way. For any given test item, the possible outcomes are 'correct','incorrect', 'abandoned' and 'not attempted'. If each individual obtains the same total score, and provided this score is below the maximum, it is possible that they have done so using quite different routes: Person A gets two correct and fails the remaining three items; Person B also gets two correct (different to those of A), gets one wrong, decides not to attempt one and abandons the remaining item. Even this diversity of pattern is oversimplified. Any correct solution could be a 'lucky guess' and one that it incorrect a 'mistake' which might be uncovered by close questioning of the subject. Eysenck (1967a) asks "Can it really be maintained that the mental processes and abilities.......are identical, merely because they obtained the same final mark?"

Thorndike and his co-workers were aware of this problem, although they discerned it in a somewhat different context, that of the measurement of speed. Because of the importance of their analysis for Furneaux's approach, the comments of Thorndike et al. (1927) are quoted in full -

> "In the measurements that are actually used it is customary to have the time a mixture of (1) the time spent in doing some tasks correctly, (2) the time spent in doing other tasks incorrectly and (3) the time spent in inspecting other tabks and deciding not to attempt them. This confusion may be permissable, or even advantageous, in the practical work of obtaining a rough measure of intellect at small expense of labour and skill, but for theory at present and for possible improvement of practice in the future we need to separate the speed of successes from the speed of failures". (p.33).

As Eysenck (1973a) notes, Thorndike never followed through the implications of his critique. Instead, it was Furneaux (1961) who attempted to incorporate these as part of his procedure for the analysis of item solutions. This involved replacing 'inspection time' by "the time spent in attempting a task and deciding to give up attempts at solution". (Eysenck 1973a, p.191).

The need for a detailed conceptual analysis of item solution as a pre-requisite of a scientific approach was a theme in the writings of both Spearman and Thorndike. This theme was taken further by Furneaux (1961). Like his predecessors, Furneaux accepted that the empirically established relationships between I.Q and other criteria justified the practice of crude testing. What remained as unsatisfactory was the incompleteness of the ensuing description of test-taking behaviour. Furneaux argued that the success of applied testing may simply be a consequence of the tests and the real-life situation reflecting an even closer relationship between only certain aspects of test-taking and real-life performance. The remaining components of the test score then simply act as a source of error. For Furneaux, the appropriate scientific goal is to minimize the error by a process of maximizing the number of categories into which test-taking behaviour can be classified and ultimately scored. These categories should be refined to the point that further sub-division is no longer possible. Hence, for Furneaux -

> "......the only really satisfactory approach to the study of test-taking behaviour is that of the thorough-going logical atomist"

While he never formulated it as such, what Furneaux appeared to be aiming at was a fractionation of the observed score variance into its component sources of variance, a model compatible with 'the analysis of variation'. The major constituents to emerge in Furneaux's model were speed, accuracy and continuance. The nature and inter-relationships of these components will be considered in later sections of this thesis.

Thus far, it has been noted that a schism between mental testing and experimental psychology developed early in the history of psychology. Although the importance of mental testing as an applied aspect of psychology has been readily accepted by its critics, they have at the same time seriously questioned its scientific status. In doing so, these critics have advanced alternative conceptualizations considered to be more consistent with a scientific approach.

Spearman's (1904b) prognostication - that the Binet-Simon style tests were "likely to have more practical than theoretical value" - appears not to have been falsified. The Binet Test has undergone several revisions (Terman and Merrill, 1960) and together with a number of tests modelled on the same pattern, continues to be used for most practical purposes in the various fields of applied psychology. The practical value of these tests cannot be denied. Their theoretical significance is questionable even though a number of test authors have made excursions into debates on the nature of intelligence (e.g Wechsler 1958). The mental test tradition, as this stream may be called, also played an important part in the development of test theory (Gulliksen 1950; Lord and Novick, 1968) and for this reason as well is of significance. Any theory of intelligence can only be tested by recourse to some form of psychological test and test theory provides the guiding principles for their construction and evaluation.

From the point of view of the experimental psychologist concerned with intelligence, the mental test tradition has provided a particular style of item and procedures for constructing and evaluating tests. Although it also provided a test format and a stimulus to use the total score, these features have served to cloud rather than clarify the task of the experimentalist.

3. The Limitations of Factor Analyses of Test Performance.

The proponents of a scientific approach have also had to contend with the second major stream in the psychology of intelligence, that encompassed by factorial analyses of intelligence tests and factorial conceptions of ability. Fron the point of view of those espousing a scientific approach, the contributions of the 'factorists' have not been given much weight (Eysenck 1967a, Furneaux 1961).

If we follow Boring's (1957) historical analysis, this second main stream of development must find its source in Darwin's 'Origin of the Species', published in 1859. Ten years later, Galton's 'Hereditary Genius' appeared followed in 1883 by 'Inquiries Into Human Faculty and its Development'. This book by Galton "has sometimes been regarded as the beginning of scientific individual psychology and of mental tests" (Boring 1957, p.483).

The major technical innovation was the invention of the technique of correlation by Galton and its refinement by Pearson. It was Pearson who also invented a technique of factor analysis, later rediscovered by Thurstone (Burt 1949). Factor analysis,according to Burt was used like other mathematical calculations "merely as aids to verifying...... hypotheses which they had already reached on broader grounds".

The philosophical conceptions of mind and intelligence, as they had evolved by the end of the 19th century, can be broadly divided into those which regarded mental processes in unitary terms, those which regarded mind as a set of faculties, and those which conceived of the intellect as being hierarchically ordered, the latter an eclectic view advocated by McDougall. Historically, and to some extent today, the faculty view finds it major proponents in North America whereas the heirarchical conception, with its lineage traceable through McDougall, Burt and Vernon is more common in this country (Vernon 1961).

Spearman's position, particularly in its early expression, was somewhat anomalous. In his early writings he was an advocate of what he called the 'theorem of intellective unity'. He had adopted the terms 'general ability' and 'general intellectual power' directly from Galton's writing on inheritance, as he did the terms 'special aptitudes' or ' special powers'. In his later years, as is generally known, he accepted more explicitly the heirarchical structure (Spearman and Jones 1950). In this sense, his views can be aligned with those who adopt a factorial conception of intelligence.

Spearman did not regard the Binet tests as scientifically respectable (Spearman 1904b), and in his research continued to employ tests of homogeneous content. In a second paper (Spearman 1904a), he also advocated the use of correlational techniques, suggesting that these might have the important theoretical consequence of being able to reveal the 'hidden underlying cause of the variations'. While both Spearman and factor analysis begin with correlations, technically they soon part company. As Burt (1949) has pointed out, Spearman only partially accepted the work of Galton and Pearson because he believed that there was a basic difference between the physical measurements used by Galton and Pearson and psychological measurements. It is possibly for this reason that he devised the rank-difference technique and the correction for attenuation. Further, as Burt states "the novel feature in his procedure consisted not in 'factor analysis' as now understood, but rather in a method which has a close affinity with so-called 'canonical analysis'". Finally, Spearman made little use of the word 'factor': it is as Burt (1949) states "scarcely mentioned".

Factor analysis has had a significant influence on the study of intelligence. This influence persists (Cattell 1971; Guilford 1967; Guilford and Hoepfner 1971) and may well be extended even further with the availability of computer based data analysis. Whether or not it can be regarded as part of experimental psychology is an open question. Boring (1957) while recognising its technological importance does not consider it appropriate to include factor analysis and the theories it generates as part of the history of experimental psychology (see p.481).

The important limitations of factor-analysis are now well known (Butcher 1968;Heim 1970) as are the limitations which emerge when attempts are made to interpret the results of factorial studies (Vernon 1961). While recent developments in factorial techniques are likely to remove some of the significant technical difficulties (Lawley and Maxwell 1963) a number of problems will remain, particularly with regard to already published studies.

Furneaux (1961) has pointed out some of the problems of interpretation that accompany factorial solutions. He suggests, for example, that apparently established group factors may well be artifacts. In examining data from Thurstone's Primary Mental Abilities, reanalysed by Eysenck (1939), Furneaux suggests that different interpretations of the data are possible. Whereas the analysis suggested a differentiation between Visuo-Spatial and Arithmetical tests, an alternative explanation could be that the differences arose because the tests defining these factors differed in the extent to which they measured speed and accuracy.

Criticisms of this type can of course be applied to all of the research on intelligence, and particularly to the multitude of studies which base themselves on factor analysis. The validity of such criticisms is however very much dependent on the empirical and theoretical status of the alternative framework. If 'speed' and 'accuracy' are psychologically, rather than only linguistically meaningful concepts, then, as Furneaux suggests, most of the factorial work "will eventually have to be repeated".

Eysenck's critique of factorial approaches is directed partly at the use of crude scores - the 'hotchpotch', at the way in which factorial approaches have become "divorced from both psychological theory and experiment", and at the failure to recognise that despite it's usefulness as a tool, the technique cannot cope with the various fundamental demands placed on it.

In presenting this historical overview of psychological approaches to intelligence, there is a risk of imposing a structure where no such structure exists. Theoreticians and research workers thrive by adopting the ideas of others so that it can be inappropriate to isolate the strains in what is essentially a hybrid. With this risk in mind, two dominant strains in the study of intelligence have been identified, the tradition of mental testing and the approach based on factor analysis. From the point of view of those particularly inclined to align themselves with a scientific approach, neither of the dominant traditions is considered to be satisfactory .

4. The 'Scientific' Approach to Intelligence Measurement.

Among the critics of both traditions, there is a lineage with several characteristics. Not only do they espouse a scientific approach but also advocate the single test item as the basic unit for the analysis of test performance. An outstanding feature of this group is the emphasis which is given to the role of 'speed' in test performance. All three characteristics are illustrated by the following statements -

> "These more recent researches have been conducted, for the most part, by the individual method of timing and have conformed to scientific testing procedure"
> (McFarland 1928).

> "....speed is so important in the intelligent act that it has seemed to us the first factor to be studied if we are ultimately to come at an understanding of the nature of intelligence".
> (Peak and Boring 1926).

Relatively few studies have been carried out within the framework of this approach, and there have been even fewer attempts to develop a substantial theoretical framework to encompass such research. Among the outstanding studies are those of Peak and Boring (1926), Sutherland (1934), Slater (1938), Tate (1948), Cane and Horn (1951), Furneaux(1961), Russell (1968) and Brierley (1969). Some attempts at theorising are apparent in the work of Thorndike (1925), Thorndike et al. (1927) and Thurstone (1937), but the most substantial analysis is that developed by Furneaux (1961). His work has also influenced a number of subsequent approaches, notably those of White (1973a+b) and Van der Ven (1971,1974) among others.

The Furneaux-Eysenck critique of the two dominant traditions in intelligence testing are important because they highlight some generally recognised problems with these approaches. As both these writers appreciate, a negative critical appraisal, by itself, is of limited scientific value. It is more helpful to be able to replace inadequate approaches, models or theories by something which is more powerful. In part as well, the relevance of their criticisms is contingent on the validity of their own theoretical framework and its empirical substrate. Furneaux's work on problem-solving described in his 1961 paper is regarded as providing not only the theoretical model but also the empirical basis for the Furneaux-Eysenck approach. Consequently, it reinforces their critical appraisal of the two main alternatives. A heavy burden is thus placed on the work of Furneaux in particular. To the extent that his findings can be shown to have some generality beyond the tests, methods, subjects and circumstances peculiar to his study, then it will indeed be necessary to reconsider current theories and approaches to intelligence. The investigations described in this study were designed in part as a test of the generality of Furneaux's approach.

The power of the Furneaux model is threefold. Firstly, it appears to overcome one of the fundamental problems in test construction, that of item-difficulty scaling. Secondly, it incorporates speed as an integral part of the problem-solving process. Thirdly, it makes conceptual provision for non-cognitive, particularly personality, influences on test scores. Before describing the model, it is necessary to examine some related problems.

5. The Problem of Difficulty.

According to standard procedures, items are considered to be of equivalent difficulty if equal percentages pass or fail the item. While this procedure may be adequate for most of applied measurement, it has one fundamental failing. It does not take account of the possibility that the 50 or other per cent who pass item 'a' may be quite distinct from the 50 per cent who pass item 'b'. Yet 'a' and 'b' would be defined as being of equivalent difficulty. Such an assumption is hard to justify. Even if there is some overlap of individuals who pass both items, the assumption of equal difficulty is still not easily justified.

Such difficulty indices are of necessity sample dependent; to obtain generalizable difficulty values necessitates representative or substantial random samples. This solution is costly and open to the effects of population structure changes which can then invalidate the indices.

Once it is recognised that equating tests for difficulty is problematic, given that the difficulties of the components are crude, subsequent interpretation of test scores becomes questionable. If a battery of tests with varying difficulties is given to a group, than, say, the interpretation of a factor analysis becomes, at the very least, a complicated exercise. If Test 'A' is 'easy', then it could involve the use of a different set of skills to those required for test 'B', a more 'difficult' test. The resulting factor structure might be quite different to the structure which could have emerged with tests of equal difficulty.

Some findings reported by McDonald (1965) illustrate some of the problems. After carrying out a principal components analysis on the test results of two groups of subjects (on the Progressive Matrices), the second component to emerge for one of the groups (the younger of the two) was identified as a difficulty factor. Using a procedure for non-linear factor analysis, McDonald found that the apparent 'difficulty' factor was a curvature component "not identifiable as such by conventional factor-analytic techniques". Apart from the questions which these findings raise for factor analysis in general, it appears that the underlying assumption that the items, with some misplacements, increase in difficulty, is not supported, at least not for the younger subjects (mean age 13.96 years).

Furneaux (1961) has suggested that the concept of item difficulty was introduced to account for the introspective observation that the "sense of effort associated with attempts to solve some problems is stronger than that associated with others".

In 1903, E.L.Thorndike confronted his contemporaries with a number of problems, the foremost being that of discovering adequate units of mental measurement.

> "Educational science needs lists of words in spelling, of examples in arithmetic,algebra and geometry...etc: so chosen that any one will be of approximately the same difficulty as any other...... The service rendered to physical science by the inch, the ounce, the ohm, the ampere.....should be duplicated in mental science....Until we have such units all our investigations rest on insecure foundations". (p.169-170).

In 1903, Thorndike was optimistic about the emergence of a solution to the difficulty problem, suggesting that "any trained student" who possessed ingenuity and a "knowledge of elementary statistics" would overcome the problem of scaling. Over seventy years have passed and the problem is still with us. As Angoff (1971) states -

> "There is little question that educational and psychological measurement would be vastly improved if its scales could be expressed in terms that everyone would agree represent equal units of ability".

Contemporary test theorists know, as did their predecessors, what they are aiming at. An acceptable scale would be constructed in such a way that if a person passes an item of given difficulty, he will pass all items that are less difficult; if he fails an item of given difficulty, he will also fail any item of greater difficulty.

In his early writings, Thorndike (1903) accepted a measure of relative status as a form of measurement, recognising at the same time that it was not very satisfactory. As an interim solution, he proposed the use of equal percentages passing as an approximate unit. This proposal was followed up in his later works (Thorndike et al. 1927), but even then, the perfect scale was not achieved. Thus, while the scale then developed (the CAVD) was "at all points more accurate than the best scales previously available", it still needed to be "improved by more extensive experimentation". (p.472).

In developing the CAVD, Thorndike et al. (1927) drew a distinction between item difficulty and intellectual difficulty.

In doing so, they diverted attention from the problem of item scaling. Thus they asserted

> ".... for every theoretical and practical purpose in the measurement of intellectual difficulty, we should use collections of tasks rather than single small tasks. We ought to measure the difficulty of single tasks, but we can profitably measure <u>intellectual difficulty</u> only in the case of composites which contain enough kinds of tasks to represent a fair sampling of all intellect as it operates at that level......."(p.133)

Item difficulty was computed by solving the equation

$$M_t + A\ s.d._t$$

where M is a measure of central tendency
t is a given task
A is a factor the sign and magnitude of which are based on R, the percentage passing the item
s.d. is the variability for the group.

Thorndike's reasons for rejecting the single item as the focus of difficulty scaling included their low correlation with his criterion of intellect (the total score), their heterogeneous variance, the restricted range of intellect sampled by such items and their proneness to being influenced by transient effects and special knowledge. Single items, according to this view, measure "but a small part of intellect plus a large error" (p.117). His solution to the difficulty problem was to combine four sets of common content items (10 each from C - Sentence Completion, A - Arithmetic, V-Vocabulary and D-Comprehension) to create a composite. Difficulty was then determined by the percentage of his criterion group who passed the composite. For an individual to pass the composite, he had to pass 50 per cent of the items. The difficulty level of the composite could be changed by juggling the items. Thus, a given composite could be made up of items of widely varying 'difficulties'. A similar procedure was employed for the age level scaling of the Stanford-Binet in its various revisions (Terman and Merrill 1960). Although a different procedure was used for scaling the Wechsler tests, these too, like most tests of intelligence, rest on the common technique of scaling by means of percentage passing.

Subsequent attempts (e.g Thurstone 1928) to overcome the 'difficulty' problem have met with little success. Gulliksen (1950), in his discussion of item analysis within the framework of classical test theory has surveyed a variety of procedures for difficulty determination. None has succeeded in overcoming the common limitation of variation in item-parameters as a function of the ability level of the group. Such problems have led to a variety of alternatives including the use of multiple indices (e.g Heim 1970, Anstey 1966), but as yet no substantial advances have been made (Angoff 1971).

A more recent view of the problem of 'difficulty' has been presented in the information theory approach of Newell and Simon (1972). They describe their tasks as being "moderately difficult problems of a symbolic nature"(p.3). The time taken to solve problems is regarded both as an important aspect of difficulty as well as being an index of difficulty. However, they recognise that 'difficulty' requires reconceptualisation in the framework of their approach.

> "In constructing a theory of problem difficulty we should like to identify those aspects of task environment and the problem solver that are the major determinants of difficulty - whether measured by solution time or any of the alternative measures" (p.93).

'Difficulty' in the Newell and Simon formulation has to be viewed in terms of the interaction between the task environment and the programme of an information processing system. Task environment is conceptualised as a set of methods for problem solution together with an "executive structure" for selecting and applying these. An important determinant of difficulty is what Newell and Simon call the problem space. This is the set of possibilities for solving the problem as seen by the problem solver. The methods at the disposal of the solver are used to examine the elements of the problem space one at a time. In the simplest case, the entire problem space is searched using the methods available to the problem solver. In this instance "time to solution will be roughly proportional to the total size of the space". The "problem solving mechanism" proposed by Furneaux (1961) has similar features to that proposed by Newell and Simon. At this point however, their analysis of problem or item difficulty has no immediate implications for determining item difficulty. The suggestion that it can be roughly indexed by time taken is not novel.

An interesting but somewhat limited approach to item difficulty is presented in the work of Elithorn and his colleagues for the Perceptual Maze Test (Smith et al,-undated; Davies and Davies 1965; Elithorn et al. 1966). The test is made up of a triangular lattice with dots at a number of the intersections. The subject is required to trace a path from apex to base, passing through a given number of dots, while moving in a forward direction.

What is called the "subjective difficulty" of the maze, can be varied in four ways. These are the physical dimensions of the maze, the size of the background lattice, the number of the target dots on the lattice and the arrangement of the dots. Although subjective difficulty can be specified very precisely, the 'difficulty' of each maze appears to be based on the standard 'percentage passing' formula (Smith et al.,-undated).

An alternative approach to assessing the difficulty of each maze was proposed by Davies and Davies (1965). It is based on the idea that each maze has a large number of distinguishable pathways. The 'difficulty' is then related to the number of paths through the maximum number of dots on the solution path. The subject obtains a score based on the 'difficulty' of finding a path through the number of dots attained. This procedure differs from the original in that a graded score is possible for each maze. Elithorn simply scored for pass or fail. Davies and Davies (1965) define 'empirical difficulty' as the percentage of subjects who pass the maze. The two measures 'difficulty' and 'empirical difficulty' correlated + 0.77, increased to +0.94 if the dot saturation of the lattice and the branches at each choice point on the correct paths were included in the computation.

While the Perceptual Mazes lend themselves well to precise specification of various parameters, the procedures employed are not transferable to the types of test item found in common intelligence tests: they cannot provide a generalised solution to the difficulty problem because item structure in the usual intelligence test is not obviously reducible to the same elements.

Item analysis is concerned with selecting test items in such a way that the test will have certain specified characteristics, in particular, that the final test will have high validities and reliabilities. According to Gulliksen (1950), there are more than twenty methods of item analysis. The determination of item difficulty is a major component in each. Seen from the standpoint of conventional test design, difficulty determination and scaling procedures are important because of their impact on validity and reliability. For most practical purposes in testing, the inadequacies of such procedures have not been absolutely crucial, as witnessed by extensive applied testing. However, they have severely restricted the interpretations of test scores and have hampered research on intelligence, again mainly because interpretation is complicated .

Most discussions of 'difficulty' seem to assume that an item or problem has something which can be called 'its difficulty'. Brierley (1969) and others have questioned this -

> "The principal reservation one must have concerning the problem of difficulty scaling is that of the reality of an intrinsic item difficulty".
> (Brierley 1969).

Similar reservations have been expressed, implicitly or explicitly, by several writers (Campbell 1961,1964; Cane and Horn 1951; Heim 1970). Campbell (1961) for example has examined the determinants of item difficulty in relation to a number of factors -

(a). Extrinsic factors
- (i) the context
- (ii) familiarity of content
- (iii) non-intellectual factors

(b). Intrinsic factors
- (i) item qualities (complexity of content, abstractness, novelty).
- (ii) item layout

Campbell (1964) in a later paper also raised the question of the extent to which many items (particularly of the series type used by Furneaux) are prone to the effects of chance strategies which can then have a profound effect on the success of the subject. Taking series items as one example, she points out that a number of rules for solution ~~is~~ are available but only one is, in the view of the test constructor, correct.

Brierley (1969) cites an additional factor, that of irrelevant information introduced to complicate the item. As an instance, he suggests a letter series item in which the rule is given by alternate letters, the others merely serving to complicate the task.

Cane and Horn (1951) found that the position of an item in a test did not affect its conventional difficubty. However, the time spent on the item was related to its position. Open-ended questions led to shorter solution times than did multiple-choice items.

Dunn and Goldstein (1959) varied a variety of item features (number multiple choices per item, irrelevant cues, grammatical changes) and found that while they produced changes in conventional difficulty, validity and reliability of the items were not appreciably changed.

A number of studies by Heim and her colleagues (see Heim 1970) have examined a variety of influences on conventional difficulty as well as difficulty indexed by solution speed. For example, Heim (1955,1957) has found that successful solutions are proportional to the difficulty of the context: an easy test preceding a more difficult test led to a smaller percentage correct in the latter. Speed was also affected in that easy items would be answered more slowly if they were preceded by a hard test. However, the less intelligent subjects produced inconsistent findings.

A further issue is the extent to which 'complexity' and 'difficulty' are synonymous. Porebski (1954) suggests that they are the same but that complexity may be introduced by irrelevant features or by requiring the subject to have at his disposal other skills(e.g a 'good' short term memory). The information theory analysis of Newell and Simon (1972) presents a different conceptual analysis of the same issues.

These various assessments of the effects of contextual factors on conventional difficulty indices provide some support for the view that such indices are sensitive to context. However, they do not resolve the issue as to the independent existence of 'intrinsic difficulty'. Were it not for the work of Furneaux (1961), Elithorn and his colleagues (Jones et al. - undated) and Davies and Davies (1965), it would be possible to reject the notion of intrinsic difficulty and simply focus on conventional indices and their empirical relationships and determinants.

The work of Elithorn, and Davies and Davies does however point to three classes of difficulty, one of which appears to be 'intrinsic difficulty'. Jones et al.(undated) distinguish between subjective and empirical difficulty. As was noted previously, the subjective difficulty (complexity?) can be varied along any or all of four dimensions (physical dimensions, lattice size, number of dots and arrangement of dots). Empirical difficulty refers to conventional indices. The Davies and Davies (1965) procedure for indexing the properties of an individual maze leads to what they call calculated difficulty. This index completely describes the structural properties of a given maze. The important point is that the calculated difficulty is invariant. It depends entirely on the maze itself and not on the context or on any other factor extrinsic to the item. In this sense, it is possible to view each maze as having an intrinsic property which appears to be directly related to what is generally regarded as 'difficulty'. It is unfortunate that Davies and Davies (1965) did not attempt to examine the scalability of the calculated difficulties but simply treated them as ranks. Also, as noted earlier, conventional items do not readily lend themselves to such structural specification although it might yet be possible to treat, say, letter series items to a similar analysis. This could be accomplished by quantifying the number of alternative correct solutions and the number of elements in the item as well as the distractors. It might be possible to examine the predictive power of such an index, the goal being to demonstrate that such indices conform to the ideals of scaling described earlier in this Chapter (Angoff, 1971). Such an exercise is beyond the scope of this thesis.

Furneaux's (1961) solution of the difficulty problem is important because it circumvents the major limitations of conventional difficulty scaling. His solution will be considered in greater detail in a later chapter. At this point it is sufficient to note that his procedures are claimed to produce indices which are independent of the standardisation group, that they are based on an unambiguous measure, solution time, and that his indices enable the prediction of performance at different levels of difficulty.

6. <u>Recent approaches to the analysis of mental test scores.</u>

Dissatisfaction with classical approaches to mental test scoring has led to the development of a number of different models (Lord and Novick 1968; White 1973a+b; van der Ven 1971,1974; Iseler 1970; Lord 1974, among others). The major shift of emphasis in all these more recent approaches is away from the older deterministic models to approaches based on probability models (Lord and Novick 1968; Lord 1974). Of the newer models, those of White (1973a+b), Iseler (1970) and van der Ven (1974) have been particularly influenced by Furneaux's conceptual analysis. None has attempted to apply Furneaux's scaling procedures.

Possibly the most widely investigated of the current approaches are those subsumed under the generic title of 'item characteristic curve theory' (icc) (Lord 1974). These models usually have two basic components, the ability of the individual and a vector of containing parameters that fully characterize the item. It is then assumed that the probability of a correct response depends only on the level of ability of the individual and on the item parameters. In its simplest form, no assumptions are made about any characteristics of the individual. More complex forms of the model have been developed by a number of writers (e.g Birnbaum 1968 - in Lord and Novick 1968). For example the 'three parameter' version of the model has as its parameters the discriminating power of the item, item difficulty and the probability of a correct answer for individuals at the lowest levels of ability.

The development of these models, has, according to Lord (1974), been held back at the empirical level because of the problem of estimating the characteristic curve of individual items. These problems appear now to have been overcome and there is accumulating evidence on the validity and usefulness of this approach. However, it appears that more research on these models is still needed (Lord 1974).

The approach adopted by Iseler (1970) has been briefly described by Eysenck (1973a, p.192). The original presentation, in German, due to both the language and mathermatics, is beyond the comprehension of the present writer!

van der Ven (1971,1974) has developed a probabilistic model for time limited tests which, conceptually, has been strongly influenced by Furneaux. However, in its most recent form (van der Ven 1974), the author has run into difficulties because of the lack of appropriate measures of what he calls "mental effort". Thus, this model is as yet incompletely specified.

All of these approaches are limited in the sense that they do not extend themselves beyond the traditional scope of psychometrics: speed is not considered as an important factor and scope for personality variation is not incorporated in the formal models. The major exception is to be found in the work of White (1973a+b).

White's model has been strongly influenced by Furneaux's conceptual analysis as well as the more recent approaches to test theory. Unlike other probabilistic models, White assumes the availability of three sets of data, whether or not an item was correct, whether or not it was abandoned, and the response time. The model also assumes that for each subject there are "three unobservable random variables", speed, accuracy, and persistence. For each problem, it is assumed that there are two "unknown parameters", difficulty level and discriminating power. In addition to these unknowns, White also introduces the concept of "effective ability", which is a function of the speed and accuracy of the subject, and of the time that has elapsed since the presentation of the item. Again, the mathematics underlying this model are beyond the competence of this writer.

White's model has been tested out on data from the present study (see White 1973b). The results appear to be encouraging although some problems remain. As a result, White (personal communication), is currently modifying the model.

7. Conclusions.

The approach to intelligence testing which developed from the Binet tests has proved to be adequate for many practical purposes. While applied testing has accommodated (more or less) to the limitations of the various procedures, the proponents of a scientific approach to intelligence have persistently expressed dissatisfaction with these procedures. In addition to the crudity of measurement several generations of critics have highlighted the problem of "difficulty", and various models have been developed in order to circumvent or solve this problem. In recent years substantial advances have been made. Among the many solutions which have been investigated, only that of Furneaux appears to be complete, although, as will be seen in a later chapter, there has been very little work on it since the 1961 paper. Of the newer approaches, that of White (1973 a+b) comes closest to the specifications of Eysenck. Like Furneaux's model, that proposed by White incorporates a speed component as well as components which give scope for integrating intelligence measurement within the compass of personality.

III. THEORIES RELATING INTELLIGENCE AND PERSONALITY

There are many psychological theories of personality but very few which have attempted to systematically incorporate intelligence (or intelligence test performance) within their respective frameworks. The two major exceptions are Eysenck (1947,1967a) and Cattell. Guilford's work has spanned both personality and intelligence, but he appears to have made no formal attempt at integration and his recent work on intelligence has been strongly criticised (Eysenck 1967a;Horn 1973). His work is considered briefly before the more detailed exposition of the views of Cattell and Eysenck.

Guilford's "Structure of Intellect' model of intelligence is essentially a classifying system which predicts that 120 functionally independent factors will be found when the appropriate tests are given. Each factor is defined on the basis of the intersection of components from each of three dimensions, made up of five mental operations, four content components, and six product components (Guilford and Hoepfner 1971). Little attention is given to speed, and the 'cube model' has been criticised by Eysenck (1967a). However, the most telling criticisms have been proposed by Horn (1973). In a series of empirical investigations, Horn found that the factor-analytic basis of Guilford's model is "not appreciably better than the support that can be provided for theories generated by random procedures". Given these comments, as well as the points already considered, no further attention is given to Guilford's work here.

In a recently published book, Cattell (1971) has attempted to draw together research on human abilities. In this section, his views on speed and its relationship to ability and personality will be discussed. Reference will also be made here to two major reviews published by Horn (1970,1972) who has been particularly important in developing Cattell's theoretical and empirical approach.

Cattell provides a list of, at this stage, tentative empirically based primary abilities. Among those factors which receive the strongest confirmation and which also have comparatively substantial variances are

U.I (4) Perceptual Speed (Identified in more than 30 studies) based on tests involving the comparison of similarity in visual material and configurations, mirror reading and dial recognition.

U.I.(5) Speed of closure (Visual Cognition, Gestalt Perception) based on 9 studies including such tests as Street Gertalt and speed of dark adaptation.

Among the "lesser, narrower, less substantiated primaries" is

U.I. 71 Motor speed.

At least five broad factors have emerged when the intercorrelations among primary factors have been further investigated in the ability realm. Cattell (1971) identified these as

Fluid General Intelligence
Crystallized General Intelligence
Power of Visualization
Retrieval Capacity or General Fluency
Cognitive Speed
(p.106)

Associated with this pattern is the theory of crystallized (gc) and fluid (gf) abilities which asserts that there are two major attributes which have a significant impact on performance on intellectual tasks. These influences operate somewhat independently (Horn 1972) throughout development and are said to represent the basic components of intelligence. Fluid intelligence is manifested primarily in tasks which are relatively uninfluenced by culture. They are either novel or overlearned. Crystallized intelligence reflects the individual's use of concepts and aids derived from the culture. Both involve the same processes of reasoning, relation perceiving, abstracting and the like but differ mainly in the extent to which they involve culture specific learning. Some differences exist between Horn and Cattell (Horn 1972). In Cattell's views, gf is regarded as more explicitly related to hereditary - phsyiological influences. gf and gc have been studied in relation to a variety of associated functions, the details of which are presented in the reviews by Cattell and Horn. For example, there is evidence that they show a differential age decline as well as being differentially influenced by damage to the nervous system. In terms of actual tests, the Matrices load on gf whereas verbal tests such as the Mill Hill Vocabulary have been found to load on gc. Cattell (1971,p.107) notes that the Furneaux Speed and Level Tests are most strongly loaded on gf.

The third of the major factors, Visualization covers a variety of performances involving spatial manipulation (e.g form-boards) and includes 'flexibility of closure' and 'perecptual speed' as correlates.

The fourth broad factor involves retrieval of material 'from memory storage', sometimes called fluency but not analogous to the primary factors of the type which are loaded by word fluency tests.

The fifth broad factor, speed of cognitive (gs) performance, poses the greatest difficulty in trying to characterise it. As Cattell points out, it has a long history, and is the factor supposedly discussed by Spearman. Its "existence" was detected in the "early and thorough" (Cattell, 1971, p.107) studies of Bernstein (1924). Parenthically, it should be noted that it is extremely hazardous to attempt to place any reliance on Bernstein's study, a hazard which Cattell consistently ignores. The quality of Bernstein's research is such as to make any conclusions derived from it exceedingly suspect.

Cattell notes that an early conception of gs placed it in the realm of personality - temperament factors and that recently, Horn has suggested that it is an index of motivational strength operating in the actual test situation. Cattell's views are somewhat complex.

The correlational evidence according to Cattell (1971, p.64) points to a number of different speeds rather than a single general speed. In addition, our conception of speed is complicated by both semantic confusion and the variety of scoring procedures used. The two major indices, number correct and time taken generally correlate positively. More substantial correlations are found between scores on timed and untimed tests provided that subjects are asked to work quickly and that they are homogeneous with respect to age (Cattell 1971, p.65). However, under speeded conditions, when intelligence and the effects of other primary factors are partialled out "two or three generalized speed measures remain". Even when such corrections are not made and the speed stress is not introduced, a tempo factor still emerges represented by the work of Rimoldi (1951) and associated with the personality factors U.I.30 (Aloof Independence) and U.I 33 (Depression-Elation).

In addition to tempo, Cattell implicates two further sources of speed differnce. The first is identified by U.I.22 (Cortertia), the characteristic level of cortical alertness of the individual and the second U.I 16 (Assertive Ego) manifest as ambition in the test situation.

In summarizing the evidence, Cattell asserts that anything that is general in cognitive speed (the cognitive speed factor of Bernstein (1924)) is temperamental or motivational in origin and is associated with U.I (22) or U.I (16). However, the U.I (22) temperamental component"... actually extends in a confusing fashion along the frontier between ability and temperament traits" (p.65). Even though it's contribution to "variance in high level abilities is quite small", Cattell suggests that it should be included in any discussion of abilities in that it has an influence on perception and executive performance. In attempting to clarify his distinction, Cattell thus posits temperamental-cognitive speed as distinct from the speed which arises through temperamental tempo, motivation level and mood. It's peculiar property is that it "appears only when ability scores are made under "speed" instructions and in scoring a timed performance" (p.65). In a later discussion of this component, Cattell maintains that it affects speed in a broad spectrum of abilities, such as numerical performance, social skills, perceptual speed and especially mechanical speeds such as writing (p.107). This component in turn makes only a minor contribution to speed in intelligence-demanding tasks, or what Cattell refers to as "power intelligence". This further type of speed is "largely an expression of the same ability as is measured in fitness and error-freeness of response". This conception of speed is similar to Spearman's view and is supported, according to Cattell by it being located within gf rather than gs in the second order factor pattern. It is the component identified by Furneaux as intellectual speed. The nature of this conception of speed is best illustrated by the following -

> "......By any reasonable perspective this simple speed factor is a distinctly broader factor even in the cognitive realm itself, than are the two intelligences. For example it operates even more obviously in mechanical and perceptual performances than in intelligence. Speed measured in successful, intelligence problem-solving is local to intelligence (being zero if a person cannot solve the problem!). If intelligence is considered speed at all, it is speed in more complex performances than those that are typically strongly loaded by gs".
>
> Cattell (1971), p.108.

Three important features emerge from this brief examination of Cattell's work. Firstly, his approach is essentially structural in that it attempts to isolate the major elements and examine their inter-relationships. Secondly, his review of relationships among the ability factors points to the possibility of two conceptions of speed, that contained in gf and that in gs at the second order. (It is unfortunate that Cattell has called gs "cognitive speed" as this factor is loaded by tasks such as 'writing speed' and 'cancellation speed'). The implications of the second feature will be considered later. Finally, Cattell attempts to include his personality factors in discussing the relationships.

The model of personality formulated by Eysenck (1967b) proposes a number of basic personality dimensions, the most important of which are extraversion (E) and neuroticism (N). His approach differs from that of Cattell in a number of ways, particularly in the level of analysis employed. Whereas Cattell is mainly concerned with traits, Eysenck's approach is essentially typological. Eysenck and Eysenck(1969), in a large scale study employing items from the work of Guilford, Cattell and Eysenck found that at the level of higher order factors E and N emerged consistently for both men and women. Bronson (1969) in her various studies of personality in children has also identified her factors with Eysenck's E and N dimensions. Brody (1972) in his recent review of various personality models has concluded that

> "....There does seem to be support for the view that the dimensions of introversion-extraversion and neuroticism seem to be present in all systematic dimensional analyses of personality..... that the measures of these dimensions of personality developed by Eysenck do in fact get at some very fundamental characteristics of individuals".
> (p.190)

There would thus seem to be little doubt that the Eysenck model, despite its competitors, provides the best current framework for an examination of personality factors in intelligence test performance. This view is further reinforced by Eysenck's (1947,1967a,1967b,1973) consistent attempts to relate these two aspects in his more general efforts to bring the psychology of intelligence into the domain of experimental psychology.

Eysenck's theory to account for the structure of individual differences in personality is based on a two component physiological model (Eysenck 1967b) of arousal activation. Eysenck proposes two mechanisms linked respectively to the reticular formation and visceral brain. Individual differences in extraversion are accounted for in terms of reticular effects on cortical arousal whereas individual differences in neuroticisms are accounted for by visceral brain influences. The hypothesis further asserts that introverts are characterised by higher arousal levels than are extraverts, the majority of people, being ambiverts, are intermediate. Individuals with low N scores are charcterised by low levels, and high N scorers, by high levels of limbic activation. These systems are postulated to function independently except on occasions when strong emotions are generated in the individual. Such emotions can then have strong arousal effects on the cortex, produced directly or through the reticular formation. According to Eysenck's hypotheses, it is not possible for there to be a condition

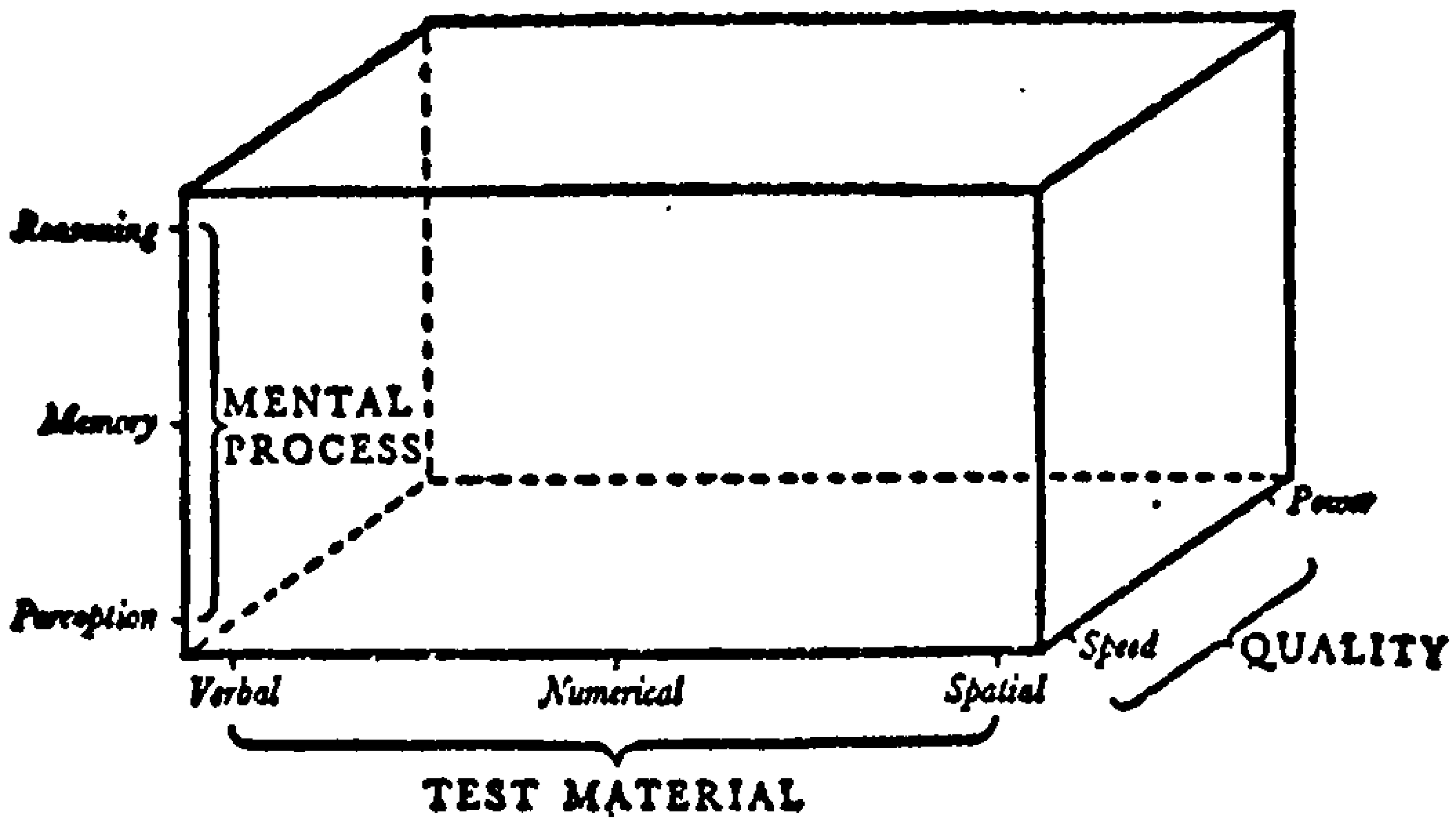


FIGURE 1. Model of the structure of intellect
Taken from Eysenck (1967a)

of strong emotion - low arousal. However, high arousal - low emotion is possible. A complete statement of the theory and both supporting and contradictory evidence are given in Eysenck (1967b).

While the basic theory summarized above appears uncomplicated, developing predictions and translating them into operations is a complex task. Eysenck (1973b) has noted some of these complexities. Predictions would ideally need to take into account the basic personality characteristics of subjects, momentary states of arousal, stimulus produced arousal, experimentally induced arousal, the role of instructions, temporal rhythms and task complexity. A further set of complications is introduced by the possibility of non-linear regression among the variables incorporated in any study. Eysenck (1967b) has asserted that non-linearty is likely to be the rule rather than the exception and that research designs need to incorporate some version of the Yerkes-Dodson law as a consequence. Yates (1973) has listed a number of predictions about personality effects on motor tasks taking some of these moderating factors into account. Eysenck (1967a) has also discussed such effects in relation to intelligence test performance.

There are a number of other aspects to the theory which are important in relation to the present study.

The Eysenck-Furneaux approach to the measurement of performance on intelligence tests has already been discussed. Eysenck has further proposed a model of the structure of the intellect which is reproduced in Fig.1. The three major dimensions encompass mental processes, test content and quality. From the diagram, it would appear that Eysenck views speed-power as a single dimension. This is not made explicit in his discussion (Eysenck 1967a) but he does assert that "mental speed and power are fundamental aspects of all mental work...." qualified to some extent by the processes and materials involved. Speed is regarded as a major source of variation, and the work of Furneaux is taken as demonstrating the fundamental nature of speed within this conception.

Regarding the theoretical relationship between speed and intelligence, Eysenck adopts Furneaux's suggestion that in problem solving, some kind of scanning mechanism functions to select a solution. The speed at which this mechanism operates determines the probability of a correct solution "being brought into focus more or less quickly". To this conception is added the notion of information processing, in the form of the hypothesis suggested by Roth's (1964) work, so that intelligence is conceived of as the speed of information processing. The rate of processing is indexed by the slope of the regression line fitted to choice reaction time data transformed into bits. This view was implicit in Furneaux's postulated mechanism but at the time he developed his theory, Roth's study was not published. As

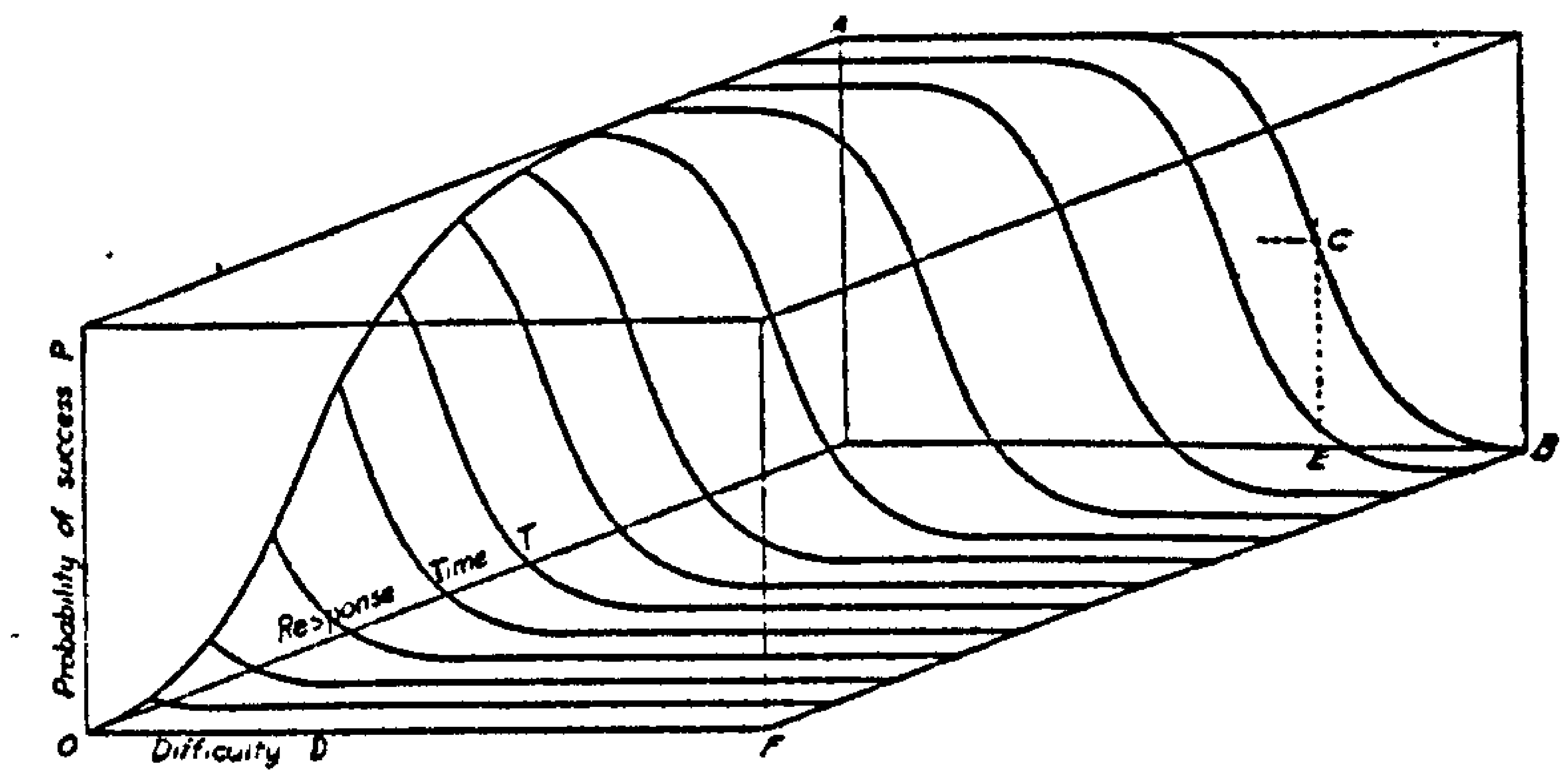


FIGURE 2. Thurstone's (1937) 'Ability Surface'
Taken from Thurstone (1937)

noted elsewhere, the data published by Roth are taken as support for Eysenck's theory.

At the empirical level speed is indexed by the mean of the log solution times to a set of 'easy' test items. This 'speed' appears not to be differentiated from 'speed' in a variety of othdr psychological tasks. In this respect, the Eysenck-Furneaux conception of speed is unlike that of Cattell. In fact, the notion of 'speed' poses a number of conceptual and operational difficulties and these will be considered shortly.

While speed receives some specification within the Eysenck model, power is less clearly formulated. Power can in practice be defined in a variety of ways, ranging from scores on untimed tests (Porebski 1960) through scores on tests which include timed components (e.g the Wechsler Tests-Wechsler 1958) to the models of Thurstone (1937) and Brierley (1969). These last two models are considered briefly in the next paragraphs before returning to Eysenck's approach.

L.Thurstone, in a theoretical paper published in 1937, examined the relationship between ability,motivation and speed, with a view to appraising "ability as power" independent of speed and motivation. As part of his model, he proposed a three-dimensional surface with difficulty defining one axis, and response time and probability of success defining the other axes.

One of the major problems, according to Thurstone, was that associated with conventional definitions of speed as the number of easy taks completed in unit time. As he saw it, the problem was whether high speed can be taken as an index of the ability to complete more difficult tasks without there being any time limits.

The three dimensional surface (see Fig.2) is generated by assuming that a subject with fixed motivation attempts a large number of tasks at each level of difficulty. All tasks are of a standard type. Their difficulty is calibrated on some scale using the percentage correct in a standard group as the index of difficulty.

From Fig.2 it can be seen that for a fixed response time, any increase in difficulty will lead to a decrease in proportion correct; as the amount of time allowed is increased, with difficulty held constant, the probability of success will increase. It will also be seen that, like Furneaux (1961), normal ogives are assumed to express relationships between response time and probability of success, and between difficulty and probability of success.

The plane AB in Fig.2 is at a point corresponding to a large valve of T, response time. The ability surface (curve) at this point has a median indicated by point C and this in turn corresponds to a difficulty level at which the probability of success is .5. According to Thurstone, if T is already a generous time allowance, any further increase in T will have relatively little effect on the "psychometric curve" ACB. On the basis of these assumptions, Thurstone is then able to propose a definition of "power or altitude which is independent of the speed of any performance".

> "The ability of an individual subject to perform a specified kind of task is the difficulty E at which the probability is $\frac{1}{2}$ that he will do the task in infinite time". (original in italics)

A practical procedure for determining E by means of interpolation is described by Thurstone. That is, various points can be determined experimentally, for example, by measuring response times at different difficulty values.

Brierley (1969) has criticized this approach on several grounds. For instance, the model disregards the effects of continuance at high levels of difficulty which is likely to produce distortions in the data. These distortions would in turn affect the adequacy of any interpolation which is attempted. As a consequence, the ability surface becomes much more complex than Thurstone realised. A further limitation of Thurstone's model is his failure to be consistent in his definition of P. At one point he refers to P as the probability "that the individual subject will successfully complete a task" and later, he refers to P as "the proportion of successful solutions".

Despite its limitations, Thurstone's model, with the exception of 'power', can fit Furneaux's theory quite closely (Brierley 1969). It gives rise to three dimensions which can be scaled (accuracy,time,difficulty) which enable an ability surface to be generated. Any given point on the surface will depend on the time limits (defined externally or internally), item difficulty, continuance, and so on. However, it seems highly unlikely that such a surface would ever be achieved in practice for any individual. It would be very difficult indeed to so manipulate practical testing so as to achieve error-free solutions at varied item difficulties, for example.

Despite the limitations of the Thurstone's model, Brierley (1969) argues for a multi-dimensional concept of power based on time, difficulty and accuracy. He suggests that if reliable and practicable units can be found for these dimensions, the generation of an ability surface is possible. Drawing on the power concept in physics, Brierley defines 'power' as the work done in unit time. It is equal to the product of the number of unit solutions achieved x difficulty x time $^{-1}$. This formulation does have a number of practical obstacles. The most important of these is that of defining a unit task. Although there has been an attempt to fractionate test items into units (Restle and Davis 1962) its outcome was suggestive rather than definite. Brierley therefore accepted that for the present, it will have to be assumed that items of approximately equivalent difficulty will have to be used. The second major problem, that of difficulty scaling and the determination of item difficulty, can be resolved by using Furneaux's (1961) procedures. Using this model of power, Brierley (1969) was able to demonstrate that diagnosed neurotics performed significantly less efficiently than did normal subjects. The details of Brierley's investigation are described elsewhere in this thesis.

The work of both Thurstone and Brierley would suggest that 'power' is a complex concept without a generally accepted definition (see Heim 1970). As noted earlier, Eysenck's conception of power does not emerge clearly in his writings.

Eysenck (1967a) in a detailed exposition of his views on test performance has also examined the relationships between learning and intelligence and learning and personality. Several studies are cited which link learning, intelligence and personality via his theory and as certain hypotheses in the present research are directly related to these, his hypotheses and the relevant evidence will be considered here.

Eysenck hypothesises that introverts should be superior to extraverts in terms of acquired knowledge. This deduction follows from two propositions, firstly that they have greater cortical arousal than extraverts and secondly, that this enhances their consolidation processes. Hence, on a task which samples such learning, for example, the Mill Hill Vocabulary Scale, introverts should achieve higher scores. Two studies (Eysenck 1947 p.124-125 and Farley, unpublished - quoted in Eysenck 1967a) produced evidence consistent with this hypothesis. However, in neither study did the effect of age appear to be considered, so that the data do not necessarily provide unequivocal support for the primary hypothesis.

Other studies of test performance and personality have also produced results consistent with various hypotheses derived from Eysenck's theory. There is some evidence that given the opportunity, extraverts tend to opt for speed whereas introverts prefer accuracy, although the evidence on the latter consists only of a non-significant trend (Eysenck 1967a).

Predictions with regard to N appear to be much more complicated but under appropriate conditions, such as stressed testing, the hypothesised curvelinear relations emerge in the data. Eysenck (1967a) also notes that if the situation is not perceived as anxiety producing individual differences in N are unimportant. In later sections of this thesis detailed consideration will be given to studies of the relationship between E,N and intelligence.

The more pervasive rôle of N in test performance has been exemplified in a study carried out by Lienert (1963),the data from which were subsequently reanalysed by Eysenck and White (1964). Lienert (1963) was able to show that children who differed in N also generate data on tests that then lead to differences in the patterns of correlation and factorial solutions. In their re-analyses of the same data, Eysenck and White (1964) were able to demonstrate a more clearly marked factorial structure in the stable than in the labile children.

In the most recent statement of his view of intelligence, Eysenck (1973) has attempted to relate the studies of cortical evoked potentials and intelligence to his theory. A number of investigations have now found significant correlations between potentials evoked by auditory and visual stimuli and intelligence, measured on a variety of tests.

Shucard and Horn (1972) correlated visual evoked potential latencies with scores on a battery of primary mental ability tests. The tests were selected to give measures of the major second order factors (G,Gf,Gc etc.). 108 subjects, of both sexes, ranging in age from 16 to 68 years were tested. The Furneaux Speed and Level tests were included in the battery. Shucard and Horn found evidence of low but reliable correlations between the latency of the average evoked potential and various measures of ability, long latencies being associated with low ability and short latencies with high ability. The correlations ranged between +.05 and -.32, the majority being around -.15 (For 100 subjects, an 'n' of .195 or greater is significant at the .05 level). In their overall analysis, the correlations between Gf and Gc with the latency measures were similar and there were significant correlations with simple cognitive processes such as perceptual motor speed.

When age was partialled out of the correlations, only those correlations between an overall measure of G,Gf and Gc remained significant. The correlations between Gf for intellectual speed (Gfsp) and level (Gflv) were reduced to non-significance (from -.21 to -.13 and from -.20 to -.13 respectively). No attempt was made to examine the data for sex differences, even though there is sufficient evidence of such differences on a variety of cognitive measures (Cooley and Lohnes 1968).

Hendrickson (1972), studying a group of 93 paid adult volunteers, used auditory evoked potentials correlated with the verbal, spatial and total scores of the AH4. Virtually all of the correlations she found were significant at the .05 level, ranging from -.3 to -.5 for latency and -.22 to -.37 for amplitude,with total score. Age effects were examined in her study but were found to have no appreciable effects on the correlations. Hendrickson has suggested that in part, the reason for the higher level of correlation is due to the more adequate measurements obtained from auditory as opposed to visual stimuli.

Eysenck (1973) has suggested that the measures of evoked potential may reflect the speed of processing information as it enters the cortex, thereby bringing this work into the theoretical framework which he has proposed for intelligence.

Among the more important problems raised by the Eysenck-Furneaux approach are those of the general validity of Furneaux's work, the empirical evidence proposed as supporting predictions from Eysenck's theory as these affect test performance, and the concept and measurement of speed. These problems are considered in the next Chapters.

IV. THE PROBLEMS OF SPEED.

1. The Conceptual Problem

The problems inherent in conventional tests led Furneaux (1961) to conclude that some other approach to intelligence testing should be devised. In effect, this involved "setting on one side the whole of the approach to cognitive function which originated with Binet" and which had "come to be taken for granted ever since". The alternative approach devised by Furneaux was based on evidence (Slater 1938; Tate 1950; Furneaux 1948) that studies of response rate were "simple,unambiguous and theoretically and practically relevant". Further, measures of response-rate also appeared to be such that they were not easily "redefined in terms of sets of simpler determinants...." These somewhat bold assertions are considered in this and the next Chapter.

The investigation of individual differences in the timing of stellar transits - the personal equation - probably represents the earliest of the systematic attempts to examine speed in human abilities (Boring 1957). Galton's interest in individual differences encompassed speed of reaction and in one form or another, psychologists have retained an interest in the speed of mental functioning up until the present day. This is witnessed by the ongoing research on speed of reaction in the elderly (Botwinick 1973), temperamental differences in speed (Eysenck 1967; Cattell 1971) and in the current research on reaction times (Laming 1968; Smith 1968).

Before 1900, much of the research on speed was focussed on individual differences in reaction time and the factors which influenced reaction speed. After 1900, there emerged a newer trend that began to focus on speed in relation to mental ability. As McFarland (1928) points out in his review of research on the role of speed in mental ability, the emphasis shifted to investigations of the relationship between "quickness" (as measured by reaction time) and "brightness", indexed by school performance and teachers judgments.

With the development of tests of intelligence, a new criterion of "brightness" became available to researchers, and it was with such criteria that relationships with reaction time were sought. This research has persisted although its focus is changed. Much of the current research on reaction time and intelligence is concerned with aging, reaction time being regarded as the major index of speed decline in the elderly (Botwinick 1973). Earlier studies were concerned mainly with younger subjects and the problems they investigated had a different theoretical orientation.

A further aspect of research on speed emerged with the advent of group tests of intelligence and their widespread use during the First World War (Boring 1957). Psychologists and testers alike became concerned with the effects of imposing time limits to facilitate data gathering and a number of studies were instigated in an attempt to evaluate the effects of time limits (May 1921 - see McFarland 1928; Ruch and Koerth 1923; Jones 1959).

Thorndike and his colleagues (1927) were responsible for one of the major theoretical analyses in which they posited various dimensions of the intellect, among which was 'speed'. Their analysis had an important impact on a number of researchers (e.g Peak and Boring 1926) both at the time of their exposition and subsequently (Furneaux 1961; Eysenck 1973a). Several other important theoretical analyses and empirical studies also appeared in the following decade (Spearman 1927; Thurstone 1938).

The last of the major foci for studies of speed probably had its origins at the beginning of this century when speed as a special ability was being investigated (Tate 1948).

The nature of the problem of speed is illustrated by Hunsicker's(1925) comments, viz:

> "The history of mental measurement shows few if any questions that have given rise to more general and persistent inquiry than has this one, the relation between rate and ability..... there is full agreement that there are individual differences in rate of work and individual differences in ability....The relationship between these two variables is the crux of the disagreement. Is there any relationship ? If any, how much? Is the quality of one trait revealed in the quality of the other ? Is rate of work any indication of mental ability ? If it is, of what significance is the fact for mental measurements?"

Although the approach to these questions has changed since they were summarised by Hunsicker (i.e. the use of factor analysis) the questions themselves have not changed. In one form or another, they have persisted through the years (Spearman 1927; Spearman and Jones 1950; Tate 1948; Lord 1956; Jones 1959; Brierley 1960,1969; Cattell 1971, among others), and despite the substantial body of research, psychologists appear still to be confronted by "this vexed question of speed" (Cattell 1971, p.64).

It might be presupposed that in talking about 'speed' psychologists have at their disposal a common basis for doing so. That is, that they have available a clear operational definition of speed that is generally accepted and one which acts as a referent for their academic discourses. It is only when such a definition is available that discussion and research are meaningful.

Spearman (1927) has stated that

> "As regards the measuring of speed, there is no great difficulty; for (with suitable arrangements) not much risk is run in inferring the duration of a person's mental processes from the time he takes to respond to the stimulus". (p.245)

This seemingly simple prescription conceals a number of major problems: it presupposes that all the mental processes between 'stimulus and response' are directed at 'problem solving' and that there is a suitable procedure for measuring these. Neither of these suppositions is acceptable, for reasons which will be presented below.

Peak and Boring (1926) have described some of the possibilities that might account for a difference in the amount of time required by two subjects to solve a test item. As they state,"the loss of time may be interstitial or it may be inherent in the intelligent act". In the interstitial case, the time difference arises because the slower subject while performing the relevant operations at the same speed as the fast subject, lost time "by irrelevant activities or by self distraction". In the alternative case, according to Peak and Boring, the time loss may "be inherent in the intelligent act, if, as far as the analysis can be pressed, it can be found that the constituents of the act occur more slowly in the poor subject than in the good subject.....Such a localisation we think of as a first step toward a solution of the problem of the nature of intelligence". On the basis of their investigation, unfortunately confined to only 5 subjects, Peak and Boring concluded in favour of the latter view. It is also unfortunate that these workers did not attempt to develop a theory to account for their findings. Until the appearance of Furneaux's(1961) analysis of problem solving, there was no significant attempt to develop such a theory.

Another of the findings of Peak and Boring, that of a significant correlation between item-solution time on a test of intelligence and a measure of reaction time, inspired a number of attempts at replication (Goodenough 1935; Farnsworth et al. 1927; Lemmon 1928), all of which failed to confirm their results. However none of these other studies recognised the most important feature of the original, namely, that the index correlated with reaction time was based on individually timed items. Instead, they correlated time to complete the test, and total scores, with their own measures of reaction time.

The Peak and Boring study was published at a time when most psychologists appeared to have a fairly clear conception, at least at the operational level, of what constituted a test of speed. The conventional speed tests required subjects to engage in repetitive activities, such as letter cancellation, detecting differences in simple shapes, adding 3 digits, and so on. (Burt 1909;McCall 1916; Highsmith 1924;Hunsicker 1925). In 1943, Cattell defined speed as the "rate of repetitive performance, where all content material is pereeptually given, through all cognitive levels". Thurstone (1937) regarded speed in terms of "the number of tasks that are completed in unit time, and these tasks are usually easy". A similar definition, emphasizing the easiness of tasks has recently been presented by Anastasi (1968), and most of the published studies on speed employ tests which conform to these prescriptions (e.g. Mangan 1959; Lord 1956; Lohnes 1966). In the majority of studies, no attempt is made to time individual items and it is implicitly assumed that inter-item time is a legitimate component of 'mental speed'. While technical difficulties in item-time measurement are no doubt important when items are answered rapidly, this is not the case when more difficult tests (such as "power" or "level" tests) are used. However, even when such tests are used, it is usually 'total time to complete the test' or 'number of items solved on a time-limit test' that provides the index of speed.

A somewhat different approach to the measurement of 'mental speed', following the influence of Peak and Boring, is also evident in the research literature. A number of investigators have used individual item times in their studies (Sutherland 1934; Slater 1938; Tate 1948; Cane and Horn 1951; Furneaux 1961; Russell 1968; Brierley 1969). These studies have also employed items of the non-repetitive type at different levels of 'difficulty', either on their own or alongside conventional speed measures. Such a divergence in trends was noted by McFarland in 1928 .

Perhaps this bifurcation was a reflection of the personalities of the researchers, extraverts opting for the conventional procedures, introverts for the individual timing approach!

The characteristics of these two distinct trends in research can be summarised in terms of differences between the timing procedures used, and the content and difficulty levels of the items. What they share is the concept and problem of 'speed'.

At an empirical level, one of the fundamental questions is the equivalence of these divergent procedures. It will be argued in the next chapter that, given the techniques for speed measurement employed in conventional speed tests, such equivalence will be difficult to establish so long as researchers adhere to certain procedures for gathering their data. Before doing so, it is necessary to consider certain conceptual problems. In doing this the writer is adopting Furneaux's (1961) approach, that of the "logical atomist" but one who is also influenced by the work of Peak and Boring.

Item solution time, test completion time or rate are gross measures. They span a sequence of events which may be different for the individuals being measured and yet lead to a conclusion that they have produced equivalent performances. A crude example will illustrate this. Individual 'A' completes an item in 10 seconds, as does individual 'B', except that the latter happened to break the point of his pencil and had to get another before he could record his response. Individual 'B' actually took only three seconds to get the solution but spent the rest of the time exchanging pencils. Are we to conclude that the speed of B is equivalent to that of A ?

While the above example can be dismissed as an instance of 'random error' in the time measurement, it is readily replaced by a more relevant psychological analysis, derived from the study of reaction times and the fine-grain analysis of certain motor acts such as tapping (Frith 1973; Spielman 1963).

In the study of reaction times, a number of investigators have been concerned to divide up the total time (T) into at least two components, the time occupied in executing the motor act and the time occupied by the 'mental events'. Birren (1964) has reported that the Movement Time in simple reactions is not appreciably altered (i.e the muscular reaction) as individuals get older. The age effects on reaction time appear to be more a consequence of the other aspects of reacting. (These findings are discussed in greater detail in the chapter devoted to reaction times). For present purpodes, it is sufficient to note the sub-division of T into Movement Time (MT) and the Reaction Time (TR)

i.e $T = MT + TR$

In his study of problem solving, Furneaux (1961) attempted to remove the equivalent of MT from the item solution times by special measures takeh before the problems were submitted to his subjects so that his time measures were in effect those equivalent to TR in the above equation. While such a refinement may seem of minor significance when dealing with events of extended duration, for example 2 or 3 minutes, MT or its equivalent may occupy a substantial proportion of T when the full sequence is 60 seconds or less.

The components of TR have been the subject of a number of conceptual analyses in studies of choice reaction time. These will also be described in greater detail in a later chapter. For present purposes, it will be sufficient to describe that presented by Smith (1968) in his review of research on choice reaction times, a paradigm which is more appropriate to test-item solution than is simple reaction time.

In addition to MT, Smith (1968) identifies four components:

1. The raw stimulus is "preprocessed" until a representation of it is formed.
2. The preprocessed stimulus is then compared with some other model in memory. On the basis of these comparisons, the stimulus is categorized.
3. An appropriate response is selected.
4. The response to be executed is programmed.

A theoretical analysis such as this is useful in helping to conceptualize the possible sequence of events although even at this simple level, there are a number of problems. For example, theoreticians are as yet uncertain of which, if any, of these stages occurs in parralel with one or more of the others or whether the events are serial. Whatever the case may be, there seems to be agreement that choice reactions can be conceptualized in terms of a number of components each spanning some period of time. This type of approach has of course been used by a number of writers. Welford (1969) for example, has outlined the difficulty in deciding which of a number of components is primarily responsible for the observed slowing in the sensory motor performance of older people. In discussing the possible sources, he suggests that similar factors may be responsible for the slowing also observed in "mental tasks". Welford (1969) has listed the component processes as including :

1. Recovery of material from memory
2. Short term retention
3. Strategies of action

On the basis of the foregoing, it would not be inappropriate to conceptualise TR as being comprised of a number of components (c) which take time to occur, viz

$$TR = C_1 + C_2 + C_3 - - C_n$$

so that, T now becomes

$$T = MT + C_1 + C_2 + C_3 + -- + C_n$$

For the present, the above 'equation' should not be regarded as an algebraic statement but as a shorthand psychological formulation.

The analysis of T can be taken at least one stage further. In problem solving, not all of the 'C' components need necessarily be involved in the mental processes doing the actual solving. Component C_1 might be 'trying to understand what the problem actually is', C_8 might be the 'checking mechanism' suggested by Furneaux (1961). For purposes of discussion, let C_4 be that component in which the 'brain is working on the problem'. Now, possibly in the same way that vigilance cannot be sustained indefinitely, or that even in such a simple task as tapping a stylus on a metal plate, there are gaps in performance (Frith 1973),

TABLE 1. An illustration of how hypothetical time components could summate to produce identical solution times, despite the components themselves being different.

	Component				Solution
Subject	C_1	C_2	$\sum_1^n W$	$\sum_1^n B$	Time
1	2	2	5	1	10
2	1	1	6	2	10
3	1	5	3	1	10

In the above table, only some of the C component times have been utilised and the W and B component times have been summed.

it is probable that there will be "gaps" or "blocks" (Bills 1931) which arise during C_4. These may arise because of something analogous to work produced fatigue, distractions, some micro-rhythm of the type described by Wolff (1967) or even the sort of mechanism which governs intermittent visual fixations (Shackel 1967). On the further assumption that we are able to discriminate and hence time the "work" and "rest/interruptions" in C_4, we would be able to express C_4 as made up of a number of components. Thus

$$C_4 = W_1 + B_1 + W_2 + B_2 + - - - W_n + B_n$$

where

W_1, W_2, etc. = are the times spent in actual work on problem solving

B_1, B_2 etc = are the interrupt times

Hence, T^1 can^2 now be re-written as

$$T = MT + C_1 + C_2 + C_3 + (W_1 + B_1 + W_2 + B_2 + \text{--} + W_n + B_n) + C_5 + \text{---} + C_n$$

The magnitude of T will also be a function of some effect of the "perceptual complexity" of the problem. To express this, it may be necessary to introduce a constant multiplier for some of the C components, or to add a constant to others. Different constants may be required for yet other C components depending on the difficulty of the item, and so on.

The main point to arise from this excursion into psychological atomism is "Which of these components times (one, some, or all) is to provide the index of speed?"

In his 1967 paper, Eysenck presented a table illustrating this question for total score. A similar table has been drawn up to illustrate the same point with reference to 'speed'. (See TABLE 1.)

The atomization of T has been confined to a general statement about an individual item time. No attempt has been made to include the additional components that might arise when the total time (TT) to complete a test is being considered. In this case TT would be comprised of T's for each item as well as the times taken between items. Also, it is possible that certain of the components of T might not be called into operation once the subject has had some practice at solving problems of a certain type.

The analysis of problem solving processes is not yet sufficiently far advanced for any pronouncement to be made on the validity of the foregoing suggestions. There does however appear to be some agreement that a multicomponent model is relevant to complex motor performances and to choice reaction times (Welford 1969; Smith 1968). There is further a suggestion that such a model is appropriate for mental tasks (Welford 1969). Furneaux's (1961) speculations about the nature of the problem-solving mechanism would also support this contention. Similarly, some of the

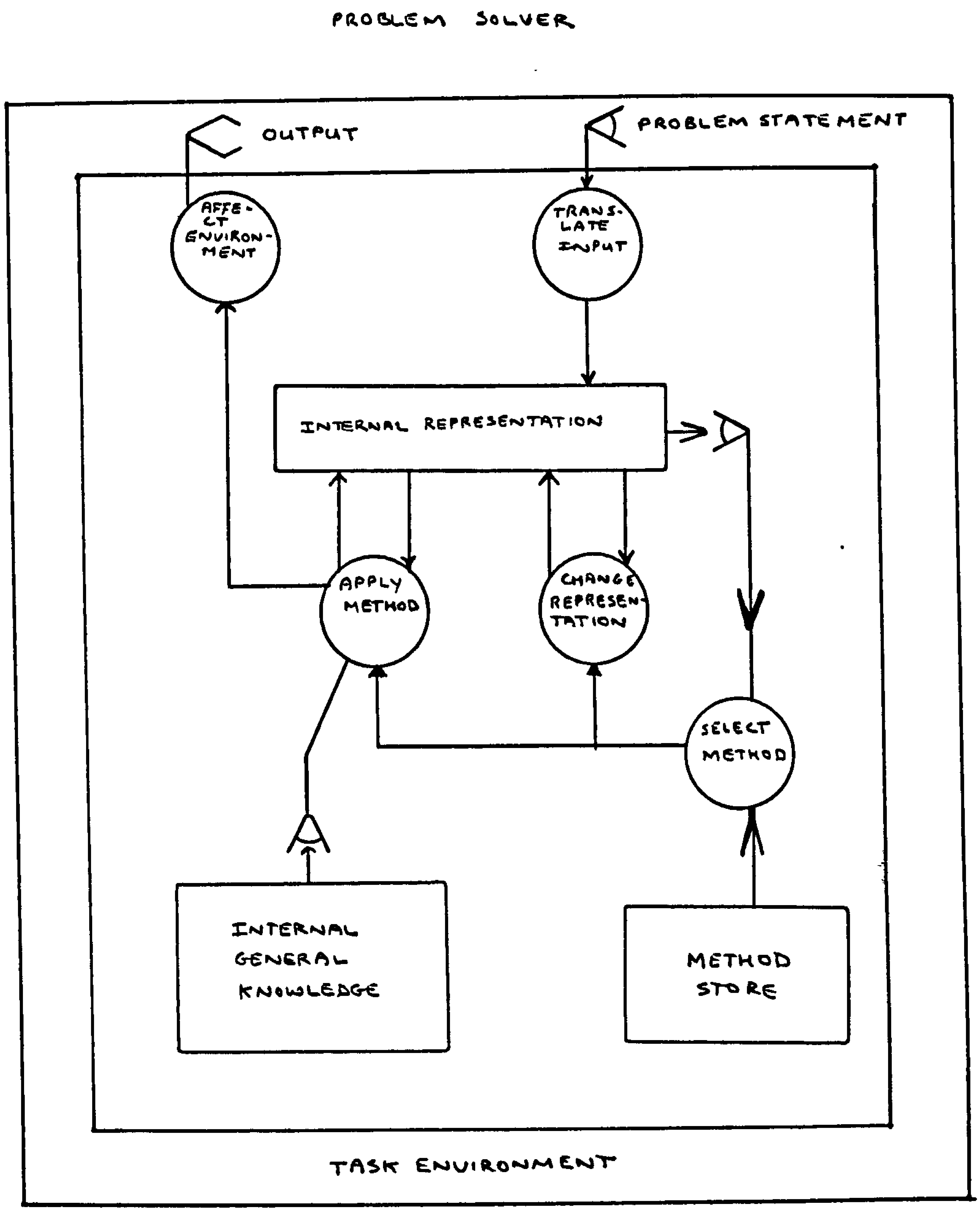


FIGURE 3. General organization of a problem solver. Taken from Newell and Simon (1972, p.89)

models described by Newell and Simon (1972) and others are consistent with such a formulation. Newell and Simon (1972) describe various information-processing systems (IPS) which involve multi-component processes, as is illustrated in Fig.3.

Thus, while a multi-component model seems to be appropriate, it is difficult to specify which of the many possible components is to be regarded as providing a basis for speed measurement. It should be emphasised that the preceding discussion has been mainly concerned with the time to solve the single item, and that even at this level, there are a number of complications in conceptualising 'speed'. These complications are extended when the unit becomes 'the number of items completed or solved' in a given period or some transformation of this index. In the final analysis however, the major constraint resides in the technology and data gathering procedures used. These are considered next.

2. The Measurement Problem

> "The inconsistencies and contradictions in the voluminous literature suggest that either speed is an unstable dimension of intellect or that inappropriate units and methods of measuring it have been frequently employed".
> (Tate 1950)

In the previous chapter an attempt was made to clarify the concept of mental speed. Such clarification is essential if suitable measurement procedures are to be devised. Tate's (1950) comments, noted above, serve to illustrate some of the problems that can arise in the absence of an appropriate procedure for measurement. Before anything can be concluded about speed, it is also necessary to have an appropriate technique for measurement.

In the view of the present writer, it is essential to have a measurement technique which ensures an isomorphism between what psychologically is supposed to be measured and what the instruments measure.

The procedures for measuring 'speed' can be classified into those which time solutions to individual items and those which record the time to complete a block of items. In the latter case, researchers then proceed to derive some index of rate, exemplified by average item solution time, total time to complete 'n' items,number of items completed in 'n'minutes,etc. Either type of index can be obtained from group or individual testing situations, and in some studies (e.g.Peak and Boring 1926), mixed procedures are used to obtain the required indices. The various combinations are presented in Table 2 with illustrative studies referenced in the cells.

In this section, it will be argued that none of these procedures provides the "proper" basis for the measurement of speed. This will be followed by a statement on what, in view of the present writer, the 'proper' procedure should be if we are to proceed to examine the role of mental speed in problem solving.

TABLE 2. Procedures for measuring speed.

TEST SITUATION	INDEX TYPE: Individual Time	INDEX TYPE: Block Time
Individual Test	Peak & Boring (1926) Brierley (1969)	Peak & Boring (1926)
Group Test	Furneaux (1961) Slater (1938)	Lord (1956) Flanagan et al. (1964)

Note. Studies repeated in the cells employed both types of procedure.

Although group-administered tests obviously facilitate an efficient collection of data, the group testing situation has a number of inherent problems which adversely influence the quality of data collected. Some of these problems are especially important in speed measurement irrespective of the type of index ultimately used in the data analysis. Other problems differentially affect individual item and 'block of items' times. Evidence for these statements is somewhat limited, mainly because authors generally fail to comment on these in the published studies. However, the nature of the difficulties can be assessed from a number of published reports (Hunsicker 1925; Tate 1950; Lord 1956; Lohnes 1966;Russell 1968; among others), as well as from more general analyses of group testing (Heim 1970;Vernon 1960;Cronbach 1970;Anastasi 1968).

Hunsicker (1925) employed both group and individual data collection procedures. In describing her study, she noted that even though her groups were in the process of being tested, they were "in the main, although not entirely free from interruptions ". She also expressed her concern with the "dishonesty" which arises in a group test setting, citing as evidence one (unreferenced) study which revealed that "fifty per cent of the class had cheated". The Hartshorne and May (1928) studies clearly indicated the severity of this problem on even simple 'speed' tests. As they noted -

> ".....even such slight changes in the situation as between crossing out A's and putting dots in squares are sufficient to alter the amount of deception both in individuals and in groups".(p.382).

This problem is not confined to children or to era. In his 1956 study, Lord pointed out that for one of his measures of speed - viz. the number of the last item attempted - there is "..reason to believe that many or all of the examinees who answered the last item of the speeded tests skipped many items or responded at random", despite being instructed not to do so.

Hunsicker (1925) also cites evidence for the unreliability of data obtained from group testing. After assessing the various procedures for collecting group data, she states

> "Not one has been found which gave evidence when in actual use, of any fair degree of control or elimination of irrelevant factors. In all likelihood, group testing by its very nature increases not only the number but the effect of disturbing elements in the situation".

After comparing her group and individual data, she states further ".....the conclusion seems beyond cavil that the group method is not dependable for securing measures of rate". As a consequence of the problems encountered, Hunsicker discarded her group data.

A more recent example of such difficulties is provided by data from Project Talent (Flanagan et al.1964;Lohnes 1966;Cooley and Lohnes 1968). In discussing the low reliabilities of the speed data, Lohnes (1966) states that these were "brought about by widespread discrepancies in the timing of the tests in different schools" (p.4-9). A similar difficulty was reported in one of the Project Talent follow-up studies (Shaycroft 1967, cited in Cooley and Lohnes 1968) where the poor stability of the speed test scores was attributed to "anomalies in retest administration" (Cooley and Lohnes 1968,p.1-16). Such difficulties arose despite the apparent sophistication of the tests and the careful plans made for their administration (Flanagan et al. 1964).

The above observations are consistent with what is known about the limitations of group tests when examiners are required to impose several short time-limits in the course of testing (Anastasi 1968). Characteristically speed tests are of short duration (Highsmith 1925;Bernstein 1924;Lord 1956; Flanagan et al.1964) and are thus particularly prone to unreliability in their administration.

The more general limitations of group test procedures have been amply documented (Anastasi 1968; Cronbach 1970;Heim 1970; Vernon 1960) and need not be detailed here. As these aforementioned authors note, they are useful for screening purposes but inadequate for precise measurement. Insofar as the measurement of speed is concerned, group tests cannot provide an appropriate basis for the measurement of performance; they may either depress relationships or distort them.

Group testing, despite its limitations, has been used to provide solution times for individual items by use of special timing devices or other procedures (Sutherland 1934;Slater 1938;Tate 1948;Cane and Horn 1951; Furneaux 1961;Russell 1968). Different techniques have been used to measure these times. Slater (1938), following Sutherland (1934) used three sets of cards numbered 0 to 9. The cards were placed on a table in such a way that one number from each set was visible to the testees. Subjects were required to record the numbers displayed when an item was completed. One card was turned by hand every two seconds so that a crude item time was measured.

Tate (1948) had items individually typed on cards. The subject wrote his answer and the time announced by the testor on each card. Furneaux (1961) employed a mechanical device which was other wise similar to that used by Slater (1938). Other timing procedures have been used, notably that by Russell (1968). In this study, a special cyclometer displayed a set of three numbers, each varying in an apparently random fashion but changing at a fixed rate. At the beginning, and after completing an item, the subject was required to look at the screen and record the number displayed. By a special decoding procedure, the time to work through a complete item could be computed.

Although such times are reported as 'solution times', this description is far from acceptable. Such times represent the duration of a sequence of activities, from turning a page, reading the problem, thinking about its answer, checking the answer and then recording it, together with any of a number of other irrelevant acts, such as correcting the solution, succumbing to distractions, changing pencils, among others. While obviously better than the gross rate measures commonly used, these times are nevertheless crude. Even using a cyclometer of the type described by Russell (1968), having to read the time adds time which is irrelevant to the problem solving process. Russell (1968) affirms "...it must be recognised that the time score is not a pure measure of the time to solution of individual items".

Brierley (1969) reports a brief investigation which he conducted into the time of irrelevant activities.In answering the Matrices, he estimates that "more than 3 minutes may well be spent simply turning pages and writing answers". It is further assumed that in these procedures the subjects will begin working on the problem immediately it comes into view. Such an assumption has not to this writer's knowledge been supported by appropriate studies.

In an attempt to overcome some of the difficulties introduced by interstitial activities, Furneaux (1961) used a special correction factor. This time constant was individually determined on the basis of subsidiary studies of the same subjects used in the main investigation, the constant being sub-tracted from the individual item times. While this is a refinement, it still does not insure accurate individual item times. Indeed, it is not possible to know how Furneaux (1961) determined such a correction factor as he does not give further details. In any case, such irrelevant activity times need to be partialled out of the data for each of the items that is to be used in reaching the answers to the research problems.

Some of the problems inherent in testing large groups can be overcome by testing small groups of four to six subjects. Hunsicker (1922) employed this procedure as did Cane and Horn (1951). These last mentioned authors devised an ingeneous apparatus which enabled them to get fairly accurate measures of individual item times. Their apparatus consisted of a kymograph with seven recording pens. One pen made a mark every 5 seconds. The remaining pens were connected to Answer Recorders, one for each of six subjects. Test items were contained in a booklet. When the subject had recorded a written answer on the answer recorder, he had to turn a knob which appeared in an apperture in the Answer Recorder. While the knob was being turned, the kymograph pen attached to his recorder vibrated. At other times, the pen was still, drawing a straight line on the kymograph. Thus, by measuring the length of the line against the 5 second marks, it was possible to get a measure item time. However, this item time still included irrelevant components, such as writing the solution, preparing to turn the knob for the next answer space etc.

While individual testing overcomes the many problems of group presentation and recording, it does not necessarily remove all of the difficulties. While the testor may be able to adjust the recorded times for some of the interstitial overt acts, it is not possible to correct for covert effects, such as knowing when the subject has actually begun his attempts at finding a solution. The Nufferno Test procedures require surreptitious timing in an attempt to overcome some of these distorting features. The present author, having used these tests in a clinical setting, is well aware of their timing limitations. Although the stopwatch is concealed, the testor has to record times on a duplicate answer sheet which of necessity, has to remain in the subjects view. Hence, the testee can become aware of the fact that each time a solution is written, the testor records some numbers on another sheet of paper. Inaccuracies also arise when the solutions are presented at a rate which the testor can't handle. Brierley (1969) has also reported similar experiences. He notes that in addition to subjects becoming aware that they are being timed, it is sometimes difficult to define precisely when the answer is written. He states "It is not uncommon for a single letter answer to be begun then left whilst the answer is re-checked before it is finally completed. Similarly a subject will often spend an appreciable period following the recording of an answer still checking that item before passing on to the next".(p.159).

Manual and electrical stopwatches, and more recently, millisecond timers are used in individual testing situations. Even under these circumstances, some researchers time blocks of items or individual as well as blocks of items (Peak and Boring 1926). Such techniques however introduce certain problems which may or may not be of any consequence. Some of these problems have been highlighted in the previous paragraphs. A further difficulty is the possible differential effect of obvious vs. surreptitious timing, or what Furneaux (1955) has called 'stressed' and 'unstressed' speed, respectively. As these procedures have a differential effect on performance (Furneaux (1955), it is necessary to treat the different studies separately : one cannot presume that 'natural speed' has been measured if the fact of timing is obvious.

In recent years a number of investigators have used complicated techniques in order to get away from some of the more obvious defects described in the previous paragraphs. Brierley (1969) constructed a special apparatus so that very little time would intervene between successive items. By housing the timing apparatus in a separate room, by arranging for the timing to begin only when the test item was presented, and by enabling the answer to be recorded when an electrical switch was depressed, the time added to problem solving time by the apparatus was trivialized. While such apparatus-working time is virtually eliminated, this technique does not

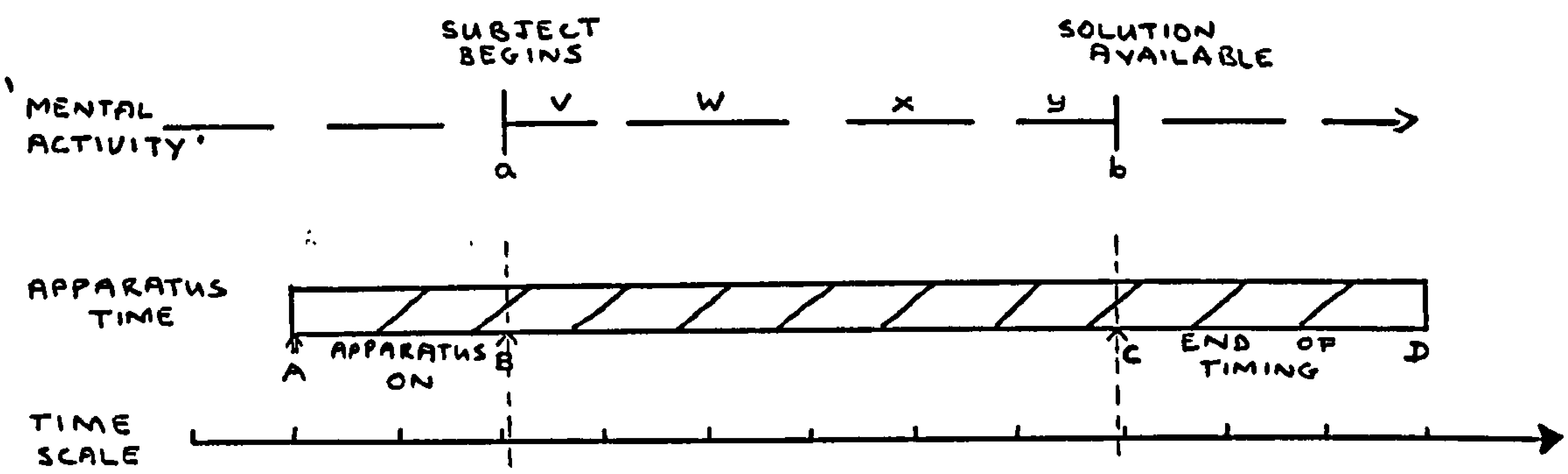


FIGURE 4. Hypothesised relationship between problem-solving activity and timing of solution.

of course remove those components of time added by interstitial activities in the subject. This problem remains however in the present study as well as in other studies.

The present writer has employed a computer-based control of item-presentation and timing to circumvent the problems of apparatus time. While the apparatus will be described in a subsequent chapter, it is worth noting at this point that timing was initiated only when the test item was exposed and terminated when the subject depressed a response button mechanically linked to a microswitch. Inter-item times were of short duration (about 0.5 sec.) and constant for all subjects. One major limitation of this approach, which only became apparent after this study was completed should also be noted here. The short delay between responding to one item and being confronted with the next means that the item is present even though the subject may not be ready or willing to begin working on it. This therefore introduces a potential interstitial time - one cannot know when the subject actually began working on the item. All that is recorded is response latency. Such a problem could be partly solved by interspersing a message which in effect tells the subject to press an 'Item Presentation Button' when he is ready for the next problem. This would not be necessary if the research was concerned with forced presentation or massed practice, although even here it would be difficult to record the time when the subject began working on the problem.

A variety of other procedures is available for the presentation of test material. These take the form of self-contained devices or else apparatus linked to computers (see papers in Elithorn and Jones 1973;Gathercole 1968; Gedye 1966; Miller 1968). However, none of these procedures is specifically designed for gathering accurate solution times.

Figure 4 is an attempt to schematize some of the foregoing discussion. The dashed line at the top represents the hypothesized periods of mental activity concerned with solving the problem. The 'true' problem solving time is the sum of components V, W, X and Y. 'a' marks the onset of the process, 'b' the point at which a solution is available. The second line represents Apparatus Time. From the point of view of the procedure, the problem could have been presented anywhere between 'A' and 'B'. Also, the end of timing could take place anywhere between 'C' and 'D', depending on what is required of the subject once he has a solution to offer. For example, if one of a set of keys has to be pushed to record a multiple choice answer, some of the time between 'C' and 'D' will be taken up in locating the correct key.

While the discussion to this point may imply an unrealistic demand for precision, it can be argued that until such time as the measurement problems are removed, research must be of limited value and any theory based on it inappropriately speculative. It is the impression of this writer that insufficient attention has been paid to the measurement problem and to an analysis of what is supposed to be measured. Yet many scientific advances are contingent on technical developments and until such time as the measurement problems are overcome, theoretical speculations can be trivial. Unfortunately, the research and theory on mental speed has failed to take such considerations into account.

The rush into speculation is nowhere stronger than in the factorial analyses of 'speed tests'. Writers have ignored the quality of their data and have proceeded to erect elaborate structures on data which are unworthy of such efforts. The problem is that if mental speed is to be investigated using factorial techniques, then the method of investigation is predetermined by the method of analysis. Proper scientific research requires the opposite. Factorial procedures need large amounts of data. The most economical way to gather such data is by group testing large numbers of subjects. Yet, as has been demonstrated in this chapter, group testing cannot, by its very nature, provide the quality of data necessary for the proper investigation of mental speed. This was apparent well before factor analysis achieved its popularity as a data analysis technique in the 1930's. It was made explicit by McFarland in his 1927 review and emphasised by Peak and Boring in their 1926 study.

If there is some substance to the multicomponent conception of speed, it is necessary to isolate the speed components for proper measurement. However, because of our current technical limitations, the basic unit of analysis must be the single item time. Such times are subject to distortions, but at least they are minimal. These times should be measured in individual rather than in group test situations as the latter make it almost impossible to cope with distorting factors.

As noted earlier, Furneaux (1961) asserted that item times (response-rates) were "simple and unambiguous" and that they could not easily be redefined in terms of simpler determinants. From the foregoing discussion, it is suggested that his assertions are unfounded. Theoretically and practically 'speed' measurement presents a number of important difficulties which are only partly overcome by timing individual items during individual testing. The further assertion that response rates cannot be defined in terms of simpler determinants is also questionable. There is sufficient evidence that response times are significantly influenced by contextual factors. Furneaux's difficulty scaling is based directly on measured response times. While difficulty may be independent of the sample, it cannot be independent of measured times.

Hence, it

cannot be independent of context. Whether or not this is an important restriction on Furneaux's solution of the difficulty problem is a task for further research, which, in addition to examining contextual factors, will also need to examine the role of instructions and other factors on performance. What evidence there is suggests that on this issue, Furneaux was over-optimistic.

3. The Role of Instructions.

There is very little research in psychology on the effects of instructions on behaviour during testing and the ways in which such instructions and their various nuances interact with other factors to produce differential outcomes. Researchers may fail to publish instructions or when reported, no evidence is given that subjects have responded to the instructions in a way which is consistant with the researcher's intentions. Attention has already been drawn to some of the ways in which the conduct of a study may contradict standardized instructions, either obviously or indirectly.

The way in which instructions are given, and the ways in which they can then be interpreted are obviously important in research on speed. Cattell (1971) has emphasised how certain "speed" factors will emerge only when "speed" is emphasised in the instructions. Eysenck (1967a) has pointed out the need to pay careful attention to structuring the test situation if the effects of N are to be manifest. Instructions are undoubtedly important here. Hence, in the design of research on speed, it is important to ensure that the instructions do not contain ambiguities and that research procedures do not contradict the researcher's intentions.

4. Concluding Remarks.

The problems inherent in measuring 'speed' are highly likely to introduce unwanted variance into solution time data. While random variance is usually coped with in the error term of analysis procedures, there is no way of knowing how much of the variance is systematic, either within subjects or across subjects. There would seem to be little doubt that "apparatus variance" is important in speed measurement. What remains unresolved is the question of apparatus x subject interaction and the extent to which it affects research on speed.

The measurement problem in speed research also includes the type of material or the task which subjects confront. As will be seen a wide variety of tasks has been used in research on speed. These range from "Making X's", through letter cancellation and adding numbers to low 'difficulty' test items, and to items from standard intelligence tests, as well as reaction times. The limitations of these tasks will be considered during the discussion of research on speed in the following chapters.

V. EMPIRICAL STUDIES OF SPEED

1. Simple and Choice Reaction Times and Intelligence.

(a). Introduction:

The study of reaction time, and the use of reaction time procedures have always been part of experimental psychology, and the study of the relationship between reaction times (as indices of speed) and intelligence is almost as old as experimental psychology itself.

Reaction time procedures entered experimental psychology following the work of astronomers and physiological psychologists on the 'personal equation'. As Boring (1957) records, the timing of various mental processes and the subtractive procedure were predominant activities in Wundt's laboratories in the 1880's. Reaction time experiments in which there is more than one signal and response were introduced by Donders in 1868 (Woodworth and Schlosberg 1955). In his classification, Donders distinguished three main procedures. The a procedure involved only the single stimulus and response, the standard RT procedure. Procedure b required one of n responses to one of n stimuli, a procedure commonly known as the CRT method. The c reaction involved a single response to one of n stimuli.

Much is now known about the factors affecting RT and CRT (Johnson 1923; Teichner 1954; Woodworth and Schlosberg 1955; Smith 1968). Both RT and CRT have always been the subject of theories aimed at explaining the underlying determinants of each and there are many attempts to account for the time difference between RT and CRT (Smith 1968). Neither of these topics will be considered in detail here except insofar as they are related to attempts to link RT and CRT with intelligence.

The association between studies of reaction times and intelligence stems in part from the common source of a concern with individual differences (Freeman 1939; Boring 1957) as well as attempts to measure such differences. Galton used a reaction time device to obtain data for his investigations of individual differences in mental ability. Galton's influence extended to James Cattell and the latter initiated an extensive programme of testing in the United States. By the last decade of the 19th century, and extending into the early part of the 20th century, many psychologists and others had devised batteries of tests which frequently incorporated reaction time procedures. The outcomes of many of the studies were reviewed by Spearman (1904b), with an essentially negative conclusion; no relationship was found between reaction time, as well as other measures, and intelligence. As Spearman put it

> "The Intellectual Order harmonizes badly with Reaction-times"

TABLE 3. Some procedural differences in the measurement of reaction time, both simple and choice.

i. Short, long, variable inter-stimulus intervals.
ii. Pre-stimulus warning, present or absent.
iii. Auditory, visual, tactile modality.
iv. Lift, press, or other response (e.g. voice switch).
v. 'Jump' response to precede responses in iv.
vi. Right, left, preferred, non-preferred limb.
vii. Number of practice trials, experimental trials.
viii. Index: mean, mode of reaction times to all or a sub-set of trials.

Whipple (1904) dismissed much of the early work as being "anthropometric" (as distinct from the scientific or "laboratory") studies not worthy of attention. He concluded "that the reaction-test is quite without significance as a measure of mental ability (save insofar as a small mean variation might indicate a certain steadiness in the control and directing of attention)".

The results of these early studies led to the gradual demise of the early tests as measures of intelligence and the loss of interest was further reinforced by the advent of the Binet test with it's items that appeared to sample the higher mental processes with greater success. As McFarland (1928) points out, after 1900 "one finds a tendency which departs from the sensory reaction time towards the emphasis on speed in relation to mental ability".

A further consequence of these early studies was the pervasive acceptance of the conclusion that reaction time is essentially unrelated to intelligence. This early response was subsequently reinforced by later studies, using the newer criterion of intelligence, which also tended to report overall trivial relationships. However, a closer examination of the literature reveals a number of studies which produce low but significant correlations between reaction time and intelligence. Before considering these, it is useful to examine the characteristics of the operations for measuring both reaction time and intelligence.

b. The measurement of reaction time.

The use of the term 'reaction time' to mean the time that has elapsed between the presentation of a stimulus and the completion of some reaction has had the unfortunate consequence of leading psychologists to assume an operational equivalence in the procedures used. As Yates (1961) has pointed out however, this assumption masks a wide range different procedures for arriving at reaction times. The extent of this diversity is illustrated in Table 3 adapted from Yates (1961). This Table is not exhaustive, nor does it list the variations in the way in which the stimuli are presented. For example, stimulus lights can vary in their intensity, the visual angle they subtend, their configuration, and so on.

The equivalence of the various procedural permutations is of course an empirical question and there is now a substantial amount of data on some of the variables which can differ from one study to the next (Woodworth and Schlosberg 1955;Thompson 1973). Procedural differences also abound in studies which measure choice times (Smith 1968).

Some of the difficulties produced by procedural variations between experiments have been described by Goldfarb (1941). In his study Goldfarb found correlations between simple, 1 of 2 and 1 of 5 choices ranging between 0.813 to 0.945 for men and from 0.894 to 0.967 for women (with age partialled out)

In his review of five studies in which simple and choice times were correlated (17 correlations) Goldfarb found a range (uncorrected) to vary between 0.27and 0.64, with a mean correlation of 0.44. Goldforb attributes these differences to differences in the procedures employed in the other studies. Whereas his subjects all reacted with the preferred hand, subjects in the other studies were required not only to discriminate between stimuli, but also to use the non-preferred hand for some of their reactions.

Given the variety of procedural possibilities, it would be inappropriate to conclude anything about reaction time as a generic concept without first conducting appropriate studies to show that such a concept is meaningful. For the present, it would seem to be more appropriate to focus on the specifics, such as procedures and subjects, in drawing conclusions. It is only when the procedural permutations have been exhausted that general statements about reaction times will be appropriate. This view is supported by some of the studies discussed later.

c. Reaction time and intelligence

Any attempt to examine the relationship between reaction time and intelligence is further complicated by the operational criterion of intelligence. Before the advent of intelligence tests as we know them today, the criterion of intelligence was either teacher judgment or some derivation of rank order according to classroom performance (which must in any case have had an influence on teacher's judgments). There was no real attempt to differentiate between speed, power or level. Once the judgment criterion was replaced by the more objective test criterion, some classification of tests was made, the usual division being between 'speed' and 'level' or 'power' tests (Goodenough 1935; Lemmon 1927; Farnsworth et al. 1927; Goldfarb 1941). However, this distinction is not the only one that can complicate an assessment of the relationship between reaction time and intelligence. Some tests use a set of homogeneous items to produce a single measure of intelligence (e.g Ravens Matrices). Others summate a variety of scores from tests sampling different abilities to produce the I.Q (e.g the Wechsler Full Scale I.Q). Still other indices of intelligence are used. Clement (1962) for example used the WAIS Vocabulary sub-test as his measure of I.Q. Not all these tests are interchangeable as I.Q measures. This is illustrated in Wechsler Sub-test scores. These sub-tests show a stable differential change with age, the Information sub-test being the least affected and Digit Symbol being most affected by aging (Botwinick 1973).

Scores on intelligence tests are derived from group and individual tests. Some of these scores are based on timed and others on untimed performance. When these basic procedural variations are them combined with the variations possible in measuring reation-time, the establishment of meaningful generalisations is a difficult task. These difficulties are further complicated by subject variables, as will be illustrated later.

d. Empirical findings using simple and choice time procedures.

No attempt will be made here to provide a detailed review of studies of simple reaction time and intelligence. There is no doubt that a substantial number of studies have found only minor relationships (Lemmon 1927;Beck 1932; Clement 1962; Farnsworth et al. 1927, and many others). There have been reviews of this relationship which have also indicated that it is of a low order (Farnsworth et al. 1927; Goldfarb 1941; Hohle 1967 among others). However, there have also been other studies which have found correlations varying in magnitude as a function of the type of test, the sex and the age of subjects (Goodenough 1935;Goldfarb 1941; Peak and Boring 1926).

Goodenough (1935), somewhat inappropriately,used a cross-sectional design in her study of the development of what she called "the reactive process". She used children aged between $3\frac{1}{2}$ and $11\frac{1}{2}$ years as subjects. The reaction time measure was a press reaction to the onset of a buzzer. Depending on the age of the child, it was given one of several tests selected from the Minnesota Pre-school Scale, the Merrill-Palmer Scale, the Arthur Performance Scale or the Stanford-Binet. All children were tested individually. Goodenough also divided her tests into those in which speed was or was not presumed to be an important factor in performance.

Correlations between 'level' tests and reaction time varied substantially, from -0.64 (7 boys,$10\frac{1}{2}$ - $11\frac{1}{2}$years) to +0.57 (12 girls aged $5\frac{1}{2}$years). The majority of the coefficients were low, and given the small numbers, within chance variations from zero. Goodenough also found that 20 of the 30 coefficients were positive, and this led her to suggest that a more extensive study would "probably reveal a very low positive relationship between reaction-speed and scores on intelligence tests that do not involve speed".

In examining the correlations between reaction times and scores on tests in which speed was considered to be an important factor,a different pattern was found. In this instance the coefficients were all positive and ranged from +0.04 (20 boys aged $8\frac{1}{2}$) to +0.58 (17 boys, $7\frac{1}{2}$ years old). Goodenough suggested that the reason for some of the lower correlations appearing in this set could be attributed to the low reliability of one of the tests. Hence, she anticipated a more general and substantial relationship with increased numbers of subjects and more adequate measures.

In the Peak and Boring (1926) study five university students were tested individually on the Army Alpha and the Otis Self-Administered Tests of Mental Ability. Subjects were instructed to be as quick and as accurate as possible and to indicate abandonments. Individual item times were observed and speed scores generated for each subject. All subjects were also given a simple reaction time task in which they had to react to the appearance of a blank dot in a circular hole in a screen. Subjects used a press response with the preferred hand and were given 110 trials. The average reaction time was computed from the last hundred trials. Peak and Boring (1926) found substantial correlations between reaction time and performance on the I.Q tests, 0.70 for the Alpha and 0.90 with the Otis. Although very cautious about the significance of their findings, due to the small number of subjects, they nevertheless were "...inclined .. to conclude that reaction time is a very important factor in intelligence as the tests test it, both because of the relation of these reaction times to the important speed-factor as analysed out of performance in the tests, and also because of their direct relation to the intelligence scores".

Although a number of researchers have subsequently attempted to repeat the Peak and Boring findings, their success has been limited and none of the correlations reached the same magnitude (Goodenough 1935;Farnsworth et al, 1927; Lemmon 1928). It should be noted however, that none of these studies used the same procedures as Peak and Boring, and they should not be accepted as appropriate tests of the Peak and Boring conclusions.

Goldfarb (1941), in his review of studies correlating reaction time and intelligence found that the 17 correlations ranged between -0.32 and +0.39, with a mean of 0. The majority of these studies were carried out on college students who Goldfarb presumed to be aged between 18 and 25 years. The purpose of Goldfarb's study was to examine the relationship between reaction time and intelligence in older subjects. His study, reported in a monograph, is too extensive to present in any detail here. However, he found statistically significant correlations between simple and choice reaction times (1 of 2, 1 of 5) for his male subjects but not for females. For his group of 108 males (age range 18 to 65 years), the correlations varied from -.131 to -.471. (The data are scored so that a positive correlation would indicate that the slower the reaction, the higher the test score). There are several sampling constraints in Goldfarb's study, such as the above average narrow range of ability in his male subjects and an inadequate sampling of female subjects at different ages. While such factors restrict the generalizability of his findings, his study does show that it is possible to observe significant correlations between intelligence and reaction time, particularly choice times. Although his correlations are low, this may have been a consequence partly of the gross intelligence measures used and the restricted range of scores for subjects on these tests.

Since the publication of Goldfarb's (1941) study, reaction time has continued to be an important variable in investigations of aging (Welford 1959; Talland 1965; Birren 1964; Botwinick 1967;1973). As Talland (1965) pointed out, "No matter how the task is defined, latency tends to increase with age". Reaction time, according to Botwinick (1973) is a "prototype of the slowing". Increased latency is observed in both simple and disjunctive measures and tends to be independent of modality and type of response required.

Birren (1964) in his review of the relationship between reaction time and age, noted that speed of reaction increases to its peak at 18 years, remains steady up to age 40, after which there is a gradual slowing of reaction. A number of experiments suggest that the main factor involved is not what he calls Movement Time, the interval needed for free muscular movement but rather the time from the appearance of a signal to the initiation of the movement. Movement times tend to change only slightly with age. By far the largest portion of the time increase is due to the reaction portion of the sequence. Birren's (1964) statements have been supported by Botwinick (1973) in his more recent review.

The increase in reaction time is, according to Botwinick (1973) partially controlled by the preparatory interval (PI) used in the reaction experiment. The PI, it is suggested, affects the state of the responder by influencing set and response preparation, with the effect on men being different to the effect on women (Botwinick 1973). With increasing age, men appear to lose their ability to maintain the set necessary for a rapid response and they also appear to need more time to recover from the effects of incorrect expectations. They also appear to need more time to prepare themselves so that they experience difficulty when responses have to be produced in quick succession. Practice reduces these effects (Botwinick 1973).

Research on choice reactions reveals that older people become disproportionately slower than younger people as the number of choices increases. This seems, according to Botwinick (1973), to be a function of how much time is allowed for the choice. If the amount of time for choice decreases, the disproportionality between old and young diminishes substantially. Some of these relationships are modified, as would be expected, by the effects of disease and organic involution (Birren 1964).

Several studies of elderly subjects have produced correlations, usually of a low order, between various measures of reaction time and a variety of cognitive tests (Birren et al. 1963; Clement 1962) although in the Birren et al. (1962) investigation, more substantial correlations have been found (ranging between 0.50 and 0.60). Clarke (1960) also reported that following an oblique rotation of the factors produced in his study, the second factor to emerge was an "unambiguous general ability factor"; choice reaction time had a loading of the same magnitude as the spatial and reasoning tests from Thurstone's Primary Mental Abilities battery, the respective loadings being .51, -.51 and -.51.

(e) Concluding remarks on reaction time studies

Experimental psychologists are concerned with producing laws which will cover relationships between variables. One of the characteristics of such laws would be that they are not context bound or bound by the methods that are used to seek these lawful relationships. Yet the overwhelming conclusion from this review of studies of reaction time and intelligence is that we have not yet come to terms with "methods variance". Substantial relationships have been reported between reaction times and intelligence. Yet these have been nullified by further studies, so that the main conclusion appears to be that there is no general conclusion. What this review of research suggests is that sex differences, age differences, test differences and criterion differences can exert an effect on outcome. There has been a tendency to regard all procedures for measuring reaction time as functionally equivalent. There has been as well a tendency to regard people, irrespective of age,sex or intelligence as equivalent. At the present time, such assumptions are inappropriate and unwarranted. If generalisations are to emerge, they will only do so once it is appreciated that methods variance appears to exert a substantial influence. It is only when this is taken into account that theory variance will become manifest.

2. Reaction Time and Personality

The research literature on reaction time has in general tended to ignore the role of personality variables. Studies may redord sex, age, modality and stimulus differences but little attention has been explicitly directed at the role of personality. For example, various reviews of reaction time, both simple and choice, do not concern themselves with the effects of individual differences in personality characteristics (Woodworth and Schlosberg 1955; Teichner 1954). Of the studies which examine reaction time in relation to 'personality', the majority have been concerned with some other problem and the personality/reaction-time relationship has been incidental (e.g Wenar 1954; Farber and Spence 1956; Stabler and Dyal 1963; Nash 1966).

The personality measure most commonly encountered is Taylor's (1953) Manifest Anxiety Scale (TMAS), used as a measure of 'drive' or motivation in studies concerned with testing predictions from Hull's (1943) learning postulates. One of the first studies to use reaction times and the TMAS in this context was carried out by Wenar (1954).

Wenar (1954) considered anxiety to correspond to what Hull (1943) called "irrelevant drive". He also used stimulus intensity as an index of "relevant" drive, basing his formulation on the established finding that in reaction time studies, increasing the intensity of the stimulus tends to lead to decreased reaction times (Woodworth and Schlosberg 1955). In his study, Wenar assumed that with other factors held constant, speed of reaction (1/RT) would be a positive function of the variations in motivation associated with variations in level of manifest anxiety and stimulus intensity. The reaction times of his subjects (84 undergraduates of unspecified sex) were measured by a key press response to a buzzer and to weak and strong shock. These stimuli constituted the stimulus intensity gradiant and subjects with scores below the 20th and above the 80th percentiles (low and high TMAS), the anxiety gradient. Wenar's results indicated that both an increase in anxiety and an increase in stimulus intensity were effective in increasing the speed of reaction during the initial training trials. However, the overall findings showed that reaction time differences between high and low anxious subjects were unaffected by changes in stimulus intensity.

This particular experiment, while indicating that high and low TMAS subjects appear to differ in speed measured by simple reaction times, also set in progress a number of other studies which, if nothing else, have added to the perplexity of the various researchers who have ventured into this field. Their confusion has not been helped by the methodological failings of Wenar's study which they have failed to take into account (e.g Faber and Spence 1956;Yates 1961) and which were pointed out by the author. As he says -

> "Since there is evidence indicating that reaction time is slower in the case of shock than in the case of auditory stimuli, the issue of how much the increase in speed of reaction in the present study was due to the increased intensity of the stimulus and how much it was due to differences in modalities is confounded.....
> Since 64 of the 84 Ss had been in an experiment in which they were shocked, and since most of them expressed concern that they would be shocked again, the entire population might be described as "shock sensitive".

Grice (1955) using the 'Air Force Discrimination Reaction Time Test'* on 30 high and 30 low TMAS basic trainee airmen found that the low-anxiety group was superior in performance on the discrimination reaction time task. However, when intellectual differences were partialled out, the superiority of the low anxious group became statistically insignificant.

In their review of the literature on the relationship between manifest anxiety and simple and choice reaction times, Farber and Spence (1956) noted that findings varied according to the type of reaction studied. One study (Wenar 1954) found a positive relationship between simple reaction time and

---

*The present author has been unable to secure a copy of the report describing this apparatus (Melton 1947).Details were not provided by Grice (1955).

manifest anxiety whereas others, using choice reaction time as the criterion, found either negative or no relationships. These findings suggested that there may be an interaction between anxiety and task complexity. Further evidence, referred to by Farber and Spence, suggested that stress factors may also interact with anxiety. Examples of such factors would include the use or threat of shock, and the intensity of the response signal used in reaction time studies.

The interrelationships among all these variables are further complicated by the manner in which TMAS is conceptualised. Two views have been noted. The first regards people with high TMAS scores as chronically anxious and thus they have high drive levels consistently. The alternative view, called the "reactive hypothesis" (Farber and Spence 1956), suggests that high TMAS scorers have a predisposition to become anxious under stress conditions. As these authors also point out, the picture is even further complicated by the manner in which manifest anxiety as a type of drive may combine with other drive variables. The research of Farber and Spence (1956) and others constitute attempts to sort out the various interrelationships and it is in this context that the association between reaction times and anxiety have been investigated.

In their first study, Farber and Spence (1956) used both simple and choice reactions to visual stimuli, varied the intensity of the stimulus and used threat of shock as a stressor. The subjects were 40 male and 40 female college students, equally divided into high and low TMAS scorers.

The data were analysed by means of variance analysis and the main effects were consistent with findings from previous reaction time studies. Men were faster than women, there were quicker reactions to the more intense stimulus, choice times (1 of 2) were slower than simple reaction times and both simple and choice times improved with practice. The p values for all of these findings were less than .001. No main effect was found for manifest anxiety or induced stress. Of the interactions found to be important, only one was of theoretical significance, mainly because it was not anticipated. This was the interaction between stress, task complexity and sex. It was found that men reacted more like women when under stress when the simple-choice time differences were examined. Farber and Spence (1956), in their attempts to understand the lack of a significant effect for anxiety, were unable to find any strong reason, although a number of possibilities were examined.

In their second study, reported in the same paper, they examined the possible role of manifest anxiety and stimulus intensity as determinants of simple reaction time. The subjects were again 40 male and 40 female students whose scores on the TMAS covered the full range. Again, the results relating to the TMAS were negative, and their overall conclusions were expressed as follows -

> "The results offered no convincing evidence that variations in amount of anxiety affected RT in any manner, either as a main effect, or as a function of stress, task complexity, stimulus intensity, or generalisation. The effect of experimentally induced stress was also unclear".

In an attempt to clarify some of the inconsistencies found in the literature on the relationship between manifest anxiety and reaction time, Kamin and Clark (1957) tested 67 normal adult males (basic trainees in the Canadian Air Force) aged 17 to 28, using an auditory RT measure. Subjects were given either non-shock blocks of trials, or trials in which they were shocked if response latency fell below a given criterion. Unfortunately, this study also failed to clarify the problem of inconsistent findings. If anything, it complicated the picture even more. Kamin and Clark found that the higher the TMAS score the slower was the simple reaction time ($r= -0.44$, p less than .01). A weaker relationship ($r= -0.24$, p less than .05) was found between the TMAS and the stress reaction time. However, it was found that the higher the TMAS score, the greater was the speed increase when subjects moved from unstressed simple reactions to the stressed reaction. These authors conclude that they were unable to account for the differences between their data and those of Farber and Spence (1956).

Stabler and Dyal (1963), taking account of Grice's (1955) findings on the importance of intelligence, carried out a study in which they also attempted to account for reaction time differences as a function of intelligence rather than of motivational factors. Their ~~group of~~ subjects, male prisoners, were divided into four groups of high and low intelligence/high and low TMAS combinations. They used visual stimuli in a choice reaction task of (1 of 3) to test their hypothesis. Intelligence was measured using the Otis tests. It was found that speed (1/reaction time) was greater for the low anxiety group on the early trials, and that as task difficulty apparently diminished with repeated practice, the performance of the high anxiety subjects improved more than that of the low anxiety subjects. The high I.Q. groups were also faster than the low I.Q. groups. The ages of these subjects were not given, nor was the possible effect of age examined in this study.

In the study of Nash et al. (1966) 36 female introductory psychology students (17.5 to 24 years), mean age 19.5 years, were placed in low, medium or high anxiety groups based on their scores on the TMAS. The subjects were then randomly divided into stress and no-stress groups, and also into groups tested by one of two experimenters. The stress condition consisted of an individually determined uncomfortable electric shock administered on 5 of the 10 reaction time trials. A simple lift reaction to light onset was used. Apart from a difference between the experimenters, the only othet difference found to be significant was that between stressed and non-stressed conditions. None of the interactions were significant. Under the stress conditions, simple reaction time was found to be slower than under the no-stress condition. Manifest anxiety appeared to have no significance in this group. The authors of this study attributed the slower reaction of their stressed subjects to the distracting effects of shock.

Costello (1968) examined the relationship between manifest anxiety, stimulus intensity and threatened (but not administered) shock on simple reaction time to a visual stimulus. Two groups of subjects, one high, the other low scorers on the TMAS were obtained from a university population. There were 30 subjects in each group, and they were randomly divided into shock-threat and no-shock-threat sub-groups.

Costello found that high anxious subjects had slower reaction times than did low anxious subjects. These findings were consistent with those of Kamin and Clark (1957) but whereas these authors had found a significant interaction between threat of shock and TMAS score, no such interaction was found by Costello.

More recently, Ferguson (1971) has found no relationship between TMAS and reaction to visual stimulus. She used 189 male and female university students as subjects but did not examine ~~his~~ her data for sex differences.

Apart from TMAS, other motivational and personality factors have been studied in relation to reaction times. Church and Camp (1965) tested 40 students on visual reaction time in order to investigate the effects of knowledge of results. All subjects were tested over a period of five days. It was found that reactions were faster if knowledge of results was provided. However, there was no evidence of a lasting effect of this information. It appeared to be effective only during the periods in which it was given.

In view of the vast amount of research generated by Eysenck's personality theory, it is somewhat surprising to encounter a paucity of research relating simple and choice reactions to these dimensions. Hence, a crucial question is whether or not the various studies using the TMAS are of any relevance?? Yates (1961,1973) appears to suggest that they are, Eysenck (1973b) writing in the same volume, questions the use of TMAS scores as measures of neuroticism and the confused results of the various studies reviewed above would suggest that Eysenck's position is the correct one. Both statements will be discussed below.

Yates (1961,1973) has based his assertions on the performance of neurotic subjects on various motor tasks. The following "relatively well supported empirical statements" (Yates 1973) provide the basis for a set of postulates, to be described shortly.

"1. Efficiency of performance on a psychomotor task is a function of drive interacting with task complexity. High drive facilitates performance when the task is of low complexity, but interferes with performance when the task is of high complexity. Drive is here defined as lability of the autonomic nervous system; complexity is defined in terms of the number of competing responses available, with the correct response being low in the response heirarchy."

Although the formality of this statement appears to give it some precision, the operational definition of complexity may well present problems. For example, failure to support this statement experimentally could always be attributed post hoc to the task not being of sufficient complexity.

"2. Neurotics* are characterized by high drive (autonomic lability) such that a stimulus that would be neutral for a non-neurotic person will trigger off a strong autonomic response in the neurotic person".

*Footnote in Yates (1973) "Actually, the reference here is to neuroticism, considered as an inherited predisposition, whereas neurosis refers to the actual state of breakdown, resulting from the interaction of neuroticism with stress"

These statements are important because they appear to suggest that data from studies with neurotics are pertinent to a discussion of neuroticism. As will be seen shortly, much of Yates' (1961;1973) discussion of the role of neuroticism is based on studies of normal subjects who vary in TMAS Scores. The link between TMAS scores and neuroticism is based on Yates' (1973) supposition that TMAS is in fact a measure of neuroticism, a point of view specifically rejected by Eysenck (1973b).

"3. Persons high in neuroticism will perform better than persons in low neuroticism where the task does not involve competing responses; the contrary will be true where competing responses are involved".

These postulates, according to Yates (1973) lead to certain predictions, which he asserts, receive empirical support.

"(a). Neurotics will show psychomotor disorganisation compared with normal subjects on complex but not simple tasks........."

"(b). An increase in the strength of the stimulus will lead to an increased amplitutde and speed in normal subjects...."

"(c) An increase in induced anxiety in normals (usually by means of some stressful situation) will facilitate amplitude and speed of simple responses......"

"(d). Whereas increasing stimulus strength, inducing anxiety or stress in normals appears on the whole to facilitate response amplitude and speed in normals where the response required is simple in nature, it is predicted from the theory that the same factors would disrupt performance where the response required had to be selected from a range of possible choices or where the response itself is of a complex nature......"

These predictions are substantially the same as those presented by Yates in his earlier (1961) review of the same topic. It is therefore surprising that the more recent statement has not incorporated the various qualifications which he presented in the earlier paper. Insofar as the research on reaction time is concerned, these postulates do not receive unambiguous support as the review in this thesis has indicated. These studies on the relationship between, TMAS, stress, stimulus intensity and simple and choice reaction times show confusing and at times conflicting results. At least in relation to reaction times, the roles of age, sex, practice and intelligence need to be clarified, as do such concepts as complexity and stimulus intensity, before predictions can be adequately stated and tested.

It is possible that another of the major difficulties with this research has been an inadequate understanding of what is being measured by the TMAS. This question has recently been discussed by Eysenck. According to Eysenck (1973b) many "investigations have shown that this scale correlates in the neighbourhood of 0.6 to 0.7 with neuroticism, and 0.3 to 0.4 with introversion..." The TMAS,will thus pick out subjects who may be introverted or neurotic (non-clinical) or both. It is however unlikely that this confusion is the sole cause of the discordant outcomes in reaction time studies.

The task of deriving testable predictions on the basis of the theory of introversion - extraversion is also complex, according to Yates (1973). While there appears to be sufficient evidence that the <u>type</u> of response for introverts will be different from that of extraverts ("provided that the experimental conditions are appropriate" - Yates 1973), other factors are involved. A task such as tapping, while apparently simple, turns out on fine grain analysis to be made up of a number of components, each of which would need to be covered in a prediction exercise. Simply measuring overall tapping rate could obscure the outcome or lead to negative results without actually affecting the theory. Further, contrasting introverts with extraverts may lead to confused data if the groups differed in neuroticism, particularly if neuroticism can influence performance. These, and other considerations have led Yates to conclude that it -

> "....is unfortunately true that detailed theoretical analysis and empirical investigations in relation to psychomotor performance have not yet been carried out so that a more comprehensive model of individual differences in psychomotor performance as related to personality factors may be constructed".

There have been a few studies which have investigated the relationship between E and N and reaction times.

Mangan and Farmer (1967) sought to relate Eysenck's theory to Russian research on psychological functioning. In this particular study, reaction times were measured in response to the onset and offset of visual stimuli of different intensities. Twenty male undergraduates, aged between 18 and 24 years were used as subjects. Unfortunately, the raw data were not analysed in relation to E and N. Instead, the authors' used a complex ratio measure which was then correlated with E and N scores from the MPI (Eysenck 1959) a scale on which E and N are in fact negatively correlated, albeit slightly. (Eysenck and Eysenck 1969;Brody 1972). However, they found that extraverts reacted near maximally at lower stimulus intensities whereas introverts did not. It is difficult to know what to make of the Mangan and Farmer conclusions as they found that E correlated 0.55 with stimulus onset, N correlated 0.47 with stimulus offset. Although both correlations were statistically significant (p less than 0.05), the latter correlation was not discussed.

In the study by Cramer (1972) no differences in response speeds were found for different groups of extreme scorers on the EPI. However, the many methodological inadequacies of this study detract from its relevance to the present discussion.

Hendrickson (1972), as part of her investigation reports that simple auditory reaction time correlated negatively with N at high but not at moderate intensities.

In Thompson's (1973) study, in which he also used simple auditory reaction times measured in 103 male subjects, some relationships emerged as statistically significant. He found some differences between introverts and extraverts in relation to the effects of the intensity of the warning signals used, the effect being more marked for the former group. N showed no relationship with reaction time.

The paucity of research on the influence of E and N on reaction times, the methodological inadequacies of some of the studies, and the many parameters which could be varied in reaction time research as well as the different components in the actual psychomotor task make it difficult to formulate any general statement. It is likely that individual differences are related to personality attributes but the necessary systematic research has yet to be undertaken.

3. Choice Reaction Time and Information Measures

Although there have been many studies of CRT, the investigation carried out by Roth (1964) appears to be the only one in which a transformation of the reaction time measures has been correlated with intelligence. In the studies cited earlier, mean reaction time, averaged over the time to respond to each set of alternatives, was the measure generally correlated with intelligence. With the advent of a set of transformations derived from Information Theory (Hick 1952; Hyman 1953) whereby the number of alternatives within a set can be expressed in terms of 'bits' of information, it became possible to relate increased choice times to amount of information by means of a linear function commonly known as Hick's Law (Smith 1968).

Before 1950, although there were relatively few CRT experiments (Laming 1968), most of the factors affecting CRT were known. In 1873, Exner demonstrated the importance of the 'preparatory set' and in 1885 Merkel showed that reaction times increased as the number of alternatives increased to ten. Age and sex differences have been observed and various factors influencing CRT have been identified (Woodworth and Schlosberg 1955).

The difference in mean reaction time between RT and CRT (the disjunctive reaction) has always been the subject of some form of theorizing (Smith 1968), but the major impetus to research on CRT itself was the publication of a series of papers which attempted to cast CRT into the framework of information theory (Hick 1950;Hick 1952;Hyman 1953). Since then, the literature on CRT has expanded rapidly and its growth continues (Smith 1968). It is not proposed to discuss in any detail the many theories and empirical findings as these are beyond the scope of the present study. The present chapter will focus only on those theories and findings which have a direct bearing on certain aspects of the information theory interpretation of CRT, one of many possible interpretations.

The characteristic CRT experiment involves at least three basic features. Firstly, through instructions given at the outset, the subject knows the alternative stimuli, responses and the associations (mapping-Smith 1968) between them. Secondly, the task is structured so that the error rates are low, usually of the order of 1% to 10%. Thirdly, no cpmparison between the stimuli is required. Finally, the major dependent variable is latency. Discussion in this section will be concerned with studies in which there is a 1:1 mapping of stimulus and response. That is, where each signal requires a different response.

In 1885 Merkel found that reaction time increased as the number of alternative signals was increased from 1 to 10 (Woodworth and Schlosberg 1955) and this finding has been repeated by many investigators (Smith 1968; Laming 1968). In the 1:1 mapping task, the mean CRT has been found to increase linearly with base 2 log of the number of alternative stimuli. This relationship between uncertainty and CRT holds when uncertainty is manipulated by varying the number of alternatives (Hick 1952) or by varying either the probabilities of occurrence of the individual stimuli or their sequential dependence Hyman (1953).

The results of many studies which have investigated the amount of information transmitted and its relationship to CRT can be summarised by the equation

$$CRT = a + b\,H_t$$

(Smith 1968)

In this equation H_t is the information measure, b is identified with processing rate and a with simple reaction time. Similar equations describing the same relationship have been presented by Bricker (1955), Schmidtke (1961) and others and it is this equation that is generally described as "Hick's Law" (Mowbray 1960).

Studies have been devised to test the immutability of Hick's Law. Mowbray and Rhoades (1959), Mowbray (1960) and Schmidtke(1961) have examined the effects of practice on CRT.

Mowbray (1960) for example suggested that the law holds only in circumstances in which the subjects are unpracticed and when the task requires the exercise of unnatural or uhfamiliar skills. In addition, he noted that many of the studies have only used a small number of subjects. In his study, 75 males and 75 females aged 17 to 48 years were required to make a verbal response to numbers displayed on read-out (Nixie) neon tubes. Each subject was given only 1 task, to respond to 1/2,1/4,1/6, 1/8,1/10 alternatives. Reaction times were analysed only to the number 8, which was common to all the tasks. The study was designed so that all the subjects had the same number of opportunities to react to the '8'. When this was done, the slope of the function was found to be 0. Mowbray (1960) concluded that CRT's to practiced responses are constant.

This conclusion was recently questioned by Burns and Moskowitz (1972) who pointed out that subjects in a 10 choice task would have five times the amount of practice as subjects in the two choice situation. "Since the critical constraint is not that each numeral appears an equal number of times but that each condition occur an equal number of times, the procedure (used by Mowbray 1960)* is faulty" (Burns and Moskowitz 1972). To overcome this problem, Burns and Moskowitz tested 20 subjects (3 females and 17 males) aged 16 to 49 years, giving an equal number of trials per presentation. They used 1,2,4,8,16 and 32 alternatives, the task being similar (naming) to that of Mowbray (1960). It was found that response latency increased as a function of the number of alternatives.

There is however evidence that the a and b constants change with

*not in original

extended practice and that given as many as 36 alternatives, the slope is curvelinear at the end of a practice session, once the number of alternatives increases beyond 12 (Schmidtke 1961).

The generality of Hick's Law has been challenged in another way. A number of researchers have suggested that the observed relationships are a function of stimulus-response compatibility (Leonard 1959; Broadbent and Gregory 1965; Fitts and Deininger 1954; Fitts and Seeger 1953; Fitts and Switzer 1962). For example, Leonard (1959) found that by using a vibratory stimulus applied independently to each of eight fingers, there was a difference between simple and two choice times but no significant difference between 2,4 or 8 choice times.

The studies just considered are but a few of many which have investigated relationships between CRT and the number of alternatives in CRT experiments. The review by Smith (1968) has looked at the conclusions of most of the studies published up until 1967 and a number of subsequent studies examined by the present writer (e.g Pachella and Fisher 1969; Hyman and Umilta 1969; Kornblum 1967, 1969; Laming 1968; Burns and Moskowitz 1972) do not appear to modify these conclusions. The major disagreements appear rather to be concerned with theoretical issues such as the validity of the information theory and other approaches.

As not all studies have been framed within an information theory approach, Smith's (1968) review has cast the empirical findings in terms of the relationship between CRT and the number of perceptually distinct choices (s). On the basis of published data, Smith states -

> "We may take as a tentative generalisation that CRT increases as the number of perceptually different choices (s) increases".

The number of stimuli in any CRT set is finite. Hence, the faster times to elements of a set could be attributed to an increase in the frequency or probability of particular stimuli, of their associated responses or both.

The evidence as assessed by Smith indicates that the more probable stimuli are responded to more rapidly and that this effect is independent of any response effects.

It has also been reliably demonstrated that reaction times to immediately repeated stimuli in a choice situation are generally faster. The evidence suggests that this effect is mainly a consequence of response repetition, although some effect can be attributed to the repetition of the stimuli as well. These findings are of importance in the design of choice experiments because they indicate a need to keep stimulus sequences constant across subjects in procedures which employ 1:1 mapping. The evidence on stimulus discriminability also carries important implications for research using CRT tasks. It has been shown that decreasing the disciminability of the stimuli, given S is held constant, leads to an increase in response times. Hence, it is important to ensure that in CRT studies not concerned with stimulus discrimination, all the stimuli are easily discriminated.

In 1:1 mapping tasks, the stimulus-response compatibility is particularly important as it can have a major impact on the b slope coefficient. Smith (1968) says "There seems to be no doubt that in such tasks, the slope of the function is inversely related to the degree of compatibility....But it is still a moot point whether this slope can be reduced to zero, that is, whether CRT is independent of S under conditions of great compatibility". The available "evidence for obtaining no relationship between CRT and S in a highly compatible task is extremely weak" (Smith 1968). The effects of practice can also reduce the b slope coefficient, although the evidence available to Smith (1968) was obtained from studies using very few subjects. However, Schmidtke's (1961) results were not considered in Smith's review, and these findings supported those obtained in the two studies reviewed.

4. Information Theory Interpretations of Choice Reaction Time

It is perhaps no great surprise that the empirical relationships demonstrated in CRT studies have led to the development of a number of theories and mathematical models which attempt to account for these relationships. All of the current theories have a common underlying logic which is similar to the logic of the earliest attempt at theorizing, contained in the work of Donders (Smith 1968). This common element is the subtraction method which in effect assumes that CRT is a complex process, the components of which can be separated conceptually and experimentally. In its present form, an important issue is whether the execution of the component stages such as stimulus categorization and responsé selection is "best approximated by a serial model or a parallel one" (Smith 1968).

The various current theories of CRT have been reviewed by Smith (1968). As he points out, the scope of contemporary theories is limited in that they attempt to account for only part of the total CRT sequence. This sequence consists of four elements, conceptualized as follows:

1. Some pre-processing of the stimulus occurs in which the stimulus is given a central nervous system representation.
2. This central representation is then compared with some already stored central representation of the alternatives and as a consequence of this comparison, the stimulus becomes categorized.

3. For this categorization, the appropriate response is selected.
4. The execution of the response is then programmed.

These, or similar processes, intervene between the stimulus onset and the emission of the response. Contemporary theories have in general not attempted to account for the fourth component and the majority, according to Smith (1968) have been mainly concerned with categorization, the second element of the sequence. Even though some models focus on a narrow aspect of CRT, there is still no generally accepted theory to account for this limited aspect. As Smith(1968) stresses, there are two fundamental issues which are still unresolved. The first is whether the central representation involved in stimulus categorization is best conceived of as "templates or lists of critical features" (Smith 1968). Secondly, it is still not possible to decide "whether stimulus categorization consists of a matching between a stimulus representation and stored templates of the possible stimuli or of a feature-testing process based on a consideration of what features the stimulus representation contains"(Smith 1968).

On the basis of these as yet unresolved theoretical problems, it would seem to be inappropriate to designate any theory of CRT as having the status of a well-supported theory. And even if the aforementioned difficulties could be resolved, there still remains the point that as yet theories have not attempted to account for the full set of processes thought to intervene between stimulus and response. As Welford (1968) has commented "Each has its advantages and difficulties, and it seems likely that there is no one model which applies in all circumstances".

Possibly the most influential of the CRT models has been that cast within the framework of communication theory, sometimes known as an 'Information Theory model'. This model has both its advocates (Hyman and Umiltá 1964) and its critics (Kornblum 1967, 1969; Laming 1968). While one of the original proponents of the information model (Hyman and Umiltá 1969) has suggested that a rejection of the information hypothesis would have been premature on the basis of the then current evidence, there were compelling grounds for at least a substantial revision of the earlier ideas. Some of the reasons are considered next.

Kornblum (1967, 1969) has presented a strong case in which he has demonstrated that the overall reaction time which is linearly related to the information measure is an artifact of the balance between stimuli which are alternations and repetitions in the stimulus sequence. The basic data show that the reaction time for a repeated event is greater than that for a non-repeated(alternative) event.

The overall mean reaction time $\overline{R}\,\overline{T}$ is the mean for the repetition $\overline{R}\,\overline{T}_r$ and that for alternations $\overline{R}\,\overline{T}_a$ each weighted by the probability of occurrence of each p_r and p_a. That is

$$\overline{RT} = P_r \overline{RT}_r + P_a \overline{RT}_a$$

(Kornblum 1967)

Komblum then goes on to argue that given $\overline{RT}_a > \overline{RT}_r$ and given a fixed number of alternatives, the "observed net increase in $\overline{RT}$ may simply be attributable to the shifts in the proportion of slower and faster responses in the overall $\overline{RT}$ distribution."

In a later study by Kornblum (1969), designed to test his views, he was able to demonstrate that the information (H) " is confounded with the probability of nonrepetition of the stimuli in most of the experimental conditions whose results have been taken as evidence of the linear relationship between choice RT and H". When these variables are unconfounded, in the experimental situation, the ensuing results lead to a rejection of the information hypothesis.

Laming (1968) has put forward arguments, on both theoretical and practical grounds, rejecting an information theory interpretation of choice reaction times. He points out that what had previously been termed the "Information Theory Model" is more appropriately called a 'Communication Theory' model. Its major theoretical weakness is the limited correspondence between the experimental situation ahd the ideal communication system.

In such a system, the <u>source</u> selects a particular <u>message</u> from a set of alternatives and this message is then transformed by a <u>transmitter</u>. The transformed message, the <u>signal</u>, is sent through a <u>channel</u> to the <u>receiver</u>. The receiver transforms the message into it's original format, but may not be able to do so accurately because <u>noise</u> has been added to the message between transmission and reception. The parallel with the reaction time experiment identifies the stimulus with the source, and the transmitter, channel, noise source and receiver with the human subject. In order to maximize transmission rate, the encoded signal must be transformed to match the statistical properties of the original. In the reaction time experiment, the transmitter is the stimulus source and its functioning is fixed and not capable of the flexibility needed to approximate the statistical structure of the signal sequence which is an essential element of Shannon's original formulation (Laming 1968).

Further ramifications of Shannons theorem also fail to be met in the communication theory analogy adopted by psychologists. Laming (and Cronbach (1955)) criticises the simple way in which the information measure (bits) of entropy has been applied. Thus, in the original statement, the measure applied only to "ideal channels in which messages are infinitely long" (Laming 1968). In the choice reaction experiment signals are of necessity very short, so that the analogy does not fit the specification in this instance as well. As a further illustration of the lack of fit between the psychological and the original model, Laming (1968) cites data from various studies of variations in the transmission rate of signals which only show a trivial/inadequate correspondence to the performance of the ideal system. As specified in the original model, the ideal communication system is capable of extended delays in responding, and indeed, such delays are considered to be an advantage in that a greater number of coding systems becomes available. Evidence from human subjects indicates however that they make progressively more errors as the delay is increased (Laming 1968). Perhaps the most telling point is that "at a theoretical level, the analogy with a communications system has not been wholehearted : rather, certain convenient aspects only of the analogy have been selected" (Laming 1968,p2).

The original studies of Hick (1952) and Hyman (1953) stimulated a great deal of research on reaction times in the context of the communication and other models. These studies were reviewed by Welford (1960,1968), Laming (1968) and Smith (1968). These reviews examined both the mathematical statements such as Hick's Law and other models not specifically cast in the framework of communication theory. From the point of view of the present study, only certain aspects are considered to be important and these are discussed in the following paragraphs.

Possibly the most important issue concerns the relationship between reaction time and entropy. The equation

$$t=a-b\sum_{i} P_i \log P_i$$

has been viewed as the fundamental choice reaction time equation (a and b are constants, P_i is the probability of signal i being presented). The entropy of the i'th signal is $-\log P_i$, so that, according to Laming (1968), "the mean reaction time to signal i should be

$$t_i = a - b \log P_i$$"

(p.10)

There is some evidence that this relationship holds when the mean reaction time is averaged over all of the signals in the series, but not when the mean time to any component signal is computed. However, in a series of investigations conducted by Laming, it was found that in certain circumstances, the series mean reaction times did not conform to the requirements of the fundamental equation. Further experiments, apart from those cited by Laming (1968) also demonstrate that choice reaction times, and hence the functions which they generate, change with practice (Schmidtke 1961), so that reaction times are more than a simple consequence of the probabilities of the signals. Another factor which influences reaction times is the discriminability of the signal, and this is not taken into account in the fundamental equation.

Studies which have investigated the channel capacity of human subjects, together with a further series of studies on the compatability of stimulus and response, also produce data which do not conform to the requirements of communication theory. For example, the model requires that the channel capacity in a true communication system is limited. Data on human subjects suggest that the channel has infinite capacity, or as Laming (1968) puts it,"that there is no channel at all" (p.14).

Laming's (1968) final comments, following his detailed analysis, are worthy of quote.

> "The theoretical anomlies and empirical weaknesses of the Communication Model are now seen to be such that the Mathematical Theory of Communication cannot provide any basis for a theory of choice reaction times". (p.16).

5. Roth's Experiment

The experiment described by Roth (1964) was an attempt to integrate an information-theory analysis of CRT within the more general compass of psychology, particularly, the study of intelligence. Drawing on the work of Hick (1952), Hyman (1953) and on the review by Schmidtke (1961), Roth examined the relationship between CRT, a test of general intelligence (the Amthauer Test) and a test which he describes as a 'pure test of speed' (the Pauli Test). As the present study was partly designed to replicate Roth's findings, his study will be considered in some detail.

The basic apparatus consisted of a stimulus generator and timing device. The stimulus generator was designed so that it would illuminate 1 light, 1 of 2, 1 of 4 or 1 of 8 lights. Associated with each light was a push button which was used by the subject to record his response. The latency of reaction, which was the interval between light and response, was recorded on the timer. The lights and response buttons were arranged so as to be equidistant from a point on which the subject rested his finger. The subject's task was to react as quickly as possible to the onset of the light. For the simple reaction time trials, all except one light and response button were covered. For the 1 of 2, only the two lights and response keys were uncovered, and so on. The subject was thus aware of the task, namely to react to the light which was on, the light being one of 1,2,4 or 8 alternatives, equivalent to 0,1,2 and 3 bits. Each subject was given 8 blocks of 20 reactions per block. Within each block, the number of alternatives was held constant. The inter-stimulus interval varied between 3.5 and 9.5 seconds. Order of presentation of the blocks was randomly varied but within each block, the interstimulus intervals remained the same for all subjects. Each block was preceded by an auditory warning signal and the experiment was carried out in a soundproof room so that subjects were isolated from any cues which might arise from the apparatus.

Roth tested a total of 85 subjects, 29 of whom were students (8 females), 18 school children (7 females) and 11 youths in a penal institution. The average age of these groups is presented in Table 4, together with the main findings of his study.

Certain features of these data are noteworthy. Firstly, Roth does not separate the data for males and females, despite the well established findings of sex differences in CRT (Woodworth and Schlosberg 1955). As they note, age for age, males are faster than females. Secondly, Roth does not report the standard deviation of the ages of his subjects. There is evidence that CRT decreases with age until age 20 years. It then tends to remain fairly stable until the mid-50's, after which CRT then shows a gradual increase (Woodworth and Schlosberg 1955).* The implications of age-related trends in CRT will be discussed shortly.

As anticipated from the research of Hick (1952), Hyman (1953) and others, Roth found a linear correlation of 0.99 between mean reaction time to each set of alternatives and number of bits of information. He computed the slope of the individual regression lines and intercepts on the ordinates for all subjects and reported finding strong inter-individual differences in these measures.

The correlations among his various measures are presented in Table 4. On the basis of these data, Roth concluded that whereas there was no relationship between IQ and simple reaction time, "the speed of information processing correlates significantly with IQ". Although Roth had hypothesised no relationship between slope and simple reaction time, this hypothesis was contradicted by the data. No attempt was made to explain this result. The measure of speed, the Pauli Test, showed no significant correlation with slope.

Some doubt on the meaning of Roth's results arises because of his inclusion of the group with a low mean IQ. Given that some individuals of low IQ may have neurological or other disabilities, it is possible that their reactions on CRT might have been impaired. The effect of this would be to artificially extend the range of reaction times, thereby increasing the likelihood of an increased correlation. Secondly, using both sexes as subjects might well have had a similar effect. Given the known sex differences in CRT, Roth should have presented evidence that for his sample, such differences were of no consequence.

*Birren (1964) states that simple times decrease up to 18 years and then remain stable until 40 years, after which they increase.

TABLE 4. Data from Roth's (1964) study.

Subjects	N	Mean Age	IQ Mean	IQ s.d.
Students (8 Female)	29	21	108	8.9
School Children (7 Female)	18	18	104	7.1
'Borstal'	11	21	85	7.3

	Correlation	t	d.f.	p
IQ and Simple Reaction Time	0.00	-	56	-
IQ and Slope	-0.39	3.15	56	< 0.01
Slope and Pauli	-0.16	0.96	33	NS
Slope and Simple R.T.	-0.41	3.34	56	< 0.01
IQ and Pauli	0.54	(not reported)		
Slope and Pauli (IQ partialled out)	0.06	-	-	-

A further limitation of Roth's data is his failure to partial out the effects of age. The relationship between age and CRT has already been noted, and given the ages of the subjects, Roth's CRT data may have been prone to age effects. That is, the correlation between slope and IQ may have been due to correlation between slope and age. In part, the strength of this limitation is also a function of the correlation between age and IQ. Although IQ's are supposed to be indepehdent of age in the population, they need not be so in a sample. In effect slope and IQ may be correlated simply because both correlate with age. Also, although in general the time and age correlation may be zero for this age range, it need not be so in a given sample. While these criticisms are speculative, they nevertheless point to serious limitations which should have been evaluated before the results of the study were used in support of his hypotheses.

Eysenck (1967a) has accepted Roth's findings without questioning the methodological adequacy of this study. He then uses these results to support his more general position, viz. that reaction time experiments can produce data consistent with a theory of intelligence based on mental speed.

In proffering this interpretation, Eysenck does not take account of two further points which can invalidate his argument. Firstly, he fails to mention Roth's finding of a non-significant correlation between slope and a measure of 'speed'. Secondly, he has not questioned the interpretation which indentifies the slope measure with speed of information processing.

The Krapelin-Pauli measure of 'speed' indexes the rate at which a subject completes a series of additions. It is not a speed measure in the sense that Furneaux defines speed, nor is it similar to other measures of speed (for example a letter-cancellation task). In fact, Roth's data suggest that the Pauli-test may well involve processes similar to those tapped by his measure of IQ. As can be seen from the data in Table 4 the Pauli and Amthauer tests produced a correlation coefficient of 0.54. Attempts to make sense of Roth's data are hampered by the lack of Pauli test data for one of his groups, and by the lack of a clear conception of what is being measured by the Pauli test. It is doubtful that it is the pure measure of speed as claimed by Roth. One author (identified as P.G in Eysenck et al.1972,p.373), describes the Pauli Test as providing "information about psychological capability (concentration,quality of performance etc.) and dynamics of performance (time taken, amount done, fatigue etc.)".

It is worth recalling at this point that correlations of a similar magnitude between choice reaction times and intelligence were reported by Goldfarb(1941). Given the near perfect correlation between choice times and the bit transformation reported in Roth's study, it is perhaps no great surprise that he found a significant correlation between slope (a function of the choice times) and intelligence.

Two fundamental questions arise out of the foregoing. The first is the empirical status of Roth's finding of a low but statistically significant correlation between slope and IQ. The second is the credibility of an Information Theory interpretation of this finding.

In the view of the present author, Roth's study is methodologically unacceptable for reasons already given. It is possible that his results are forced by the methodological failings of his experiment, so that any attempt to invest them with psychological meaning is at best premature. However, if his finding can be shown to hold in an independent study, then at least an important empirical relationship will receive additional support. The theoretical problem then becomes one of accounting for why the slope transformation of CRT data produces a correlation with intelligence. The first priority however is to produce the evidence that such a relationship holds. Thus one of the aims of this study is to test if such a relationship can be found.

The attempt to interpret Roth's finding in Information Theory terms appears at this point in time to be unwarranted. To do so requires a detailed evaluation of the status of Information Theory in relation to CRT. Some of the critical aspects of this analogy have been considered in this chapter, and the overall conclusion is that the analogy is trivial. Like any attempt to extend Roth's findings, the Information Theory interpretation is premature if not inappropriate and misleading.

6. Speed and Intelligence

(a). Introduction

This review of empirical studies on the role of speed in mental ability will take as its starting point the conclusions of McFarland's (1928) review. Although a number of reviews have been published since 1928, (Himmelweit 1946;Tate 1948;Jones 1959;Vernon 1961 ;Brierley 1960,1969; Russell 1968), none has been as detailed or as methodologically incisive as that of McFarland. For the most part, these later reviews accept the original author's conclusions without any serious attempt to discover if the conclusions are acceptable, given the methods used. On the basis of the present author's examination of the later studies, this failing of subsequent reviewers is serious, for in many cases, as will be recorded in subsequent paragraphs, the most appropriate conclusion would seem to be that no conclusion is possible due to the variety of methodological inadequacies which have characterised these studies.

The studies to be considered in this section will be divided into the following categories -

(a) Non-factor analytic group-test research

(b) factor-analytic studies

(c) studies employing individual item times

(d) studies relating speed to personality variables

The early investigations of the relationship between reaction time and intelligence were gradually supplemented by studies of the relationship between measures of rate and judgments of ability. The study by Burt(1909) in which he correlated card sorting and alphabet sorting speeds and headmasters' judgments of intelligence exemplifies the type of research carried out. This, like related studies was subsequently criticised on various grounds, particularly because of the criterion of intelligence used (McFarland 1928).

In 1916, an important conceptual complication was introduced by McCall when he raised the question of the relationship between 'speed tests', 'power tests' and intelligence. As his criterion of intelligence,McCall employed teachers' ratings, school marks and composite test scores. The speed tests included cancellation tests, addition and the power tests, vocabulary,sentence completion and other indices. This study is historically important mainly because it was probably the first to introduce the speed /power distinction (McFarland 1928).

With the advent of group tests of intelligence, and particularly the widespread use of such tests during the First World War (Boring 1957), psychologists and testees alike became concerned with the effects of time limits.(May 1921 - see McFarland 1928;Ruch and Koerth 1923).

To examine the consequences of time limits, the general procedure adopted was to administer the test in the prescribed time and then to allow extra time for those testees who needed it. Although such studies usually found substantial correlations between the limited and unlimited time scores, they tended to be methodologically inadequate. Hunsicker(1925) for example pointed out that because of the wide range in the abilities of the testees, a single test such as Army Alpha was a speed test for the more intelligent of the subjects and a power test for those who were less able. The Ruch and Koerth (1923) investigation was criticised on different grounds (Highsmith 1924) although they also found substantial correlations between limited and unlimited time administrations.

By 1925, research had become more sophisticated and a more differentiated conceptual scheme had emerged. Hunsicker (1925) commented on the need to control for accuracy, and she introduced individual item-times.

Although she strongly criticised group testing, her own study employed small groups. Also in 1925, Thorndike introduced a scheme whereby intelligence was conceptualised in terms of level or power, range and speed. By 1925 as well, the role of persistence had been proposed (McFarland 1928) and Spearman's questioning of the existence of a speed factor had entered the literature. Spearman's work will also be considered in a later section of this chapter.

It is against this background that McFarland (1928) was able to conclude that, using the method of individual timing and by using data from studies that "have conformed to scientific testing procedure",

> "....the evidence,although contradictory, decidedly tends to favour the existence of a positive relationship between rate and ability in mental tests".(McFarland 1928)

What remained unclear,according to McFarland (1928) was the amount and nature of the relationship, two issues which needed to be resolved by further research.

Whereas the majority of studies before 1930 used group-test procedures without recourse to factor analysis to analyse the data, after 1930, most of the studies relied on factor analysis. The major exception was the series of studies which examined individual item times.

b. Studies using group-testing (non-factor-analytic)

Baxter (1941) using a number of standard intelligence tests (Otis Self-Administering Test, Revised Army Alpha) as well as a number of outside criteria (examination grades), examined the inter-relationships between speed, power and level among 100 university undergraduates.

In his first study, Baxter asked subjects to work as quickly and as accurately as possible while being individually tested on the Otis. He recorded the time to complete the test (speed), the number of items correct in unlimited time (level) and the number of items completed in 20 minutes (power). These procedures were later adapted to a group presentation of a parallel form of the Otis and the inter-relationships re-examined under these conditions.

Baxter found that speed and level varied independently and that all the power score variance could be accounted for by speed and level, the contribution of speed being somewhat greater. It was also found that the inter-relationships varied as a function of the test of intelligence used and that group testing led to a reduction in the strength of the relationships.

Jones'(1959) review was concerned specifically with speed changes in relation to age. It is an important assessment of the problem in that it highlights certain variables that may differentially affect performance. From the studies considered, it was apparent that time limits can reduce the scores of elderly subjects but that after about 60 years, there is a suggestion that such differences are no longer substantial. However, the extent of any difference is a partial function of the nature of the test. Another important influence in the relationships between speed and aging is the complexity of the task. While these general conclusions are important because they direct us to variables to be controlled, none of the studies mentioned in Jones'(1959) review used individual item times. Rather, they employed indices of speed based on 'number of items attempted' (Jones 1959).

Knapp (1960) tested two groups of 100 subjects each on what he described as 'speed' and 'power' versions of the Cattell Culture Free Intelligence Test. The imposition of time limits was used to define the speed version of this test. Power and speed tests were given in a counter-balanced order in a group-testing situation, the one condition immediately following the next. Subjects in one of the groups were applicants for Visa's to enter the U.S.A and testing was part of the procedure to which these people were subjected. The only result of importance was the finding that the order of speed-power test administration was important. Initial testing on a speed test seemingly placed all subjects at a disadvantage relative to their performance in a power-first administration.

c. Factor-analytic studies of speed.

Factor-analysis has had a profound influence on our conceptions of human ability. As Horn (1972) points out, its history "is very closely tied to the history of our ideas about human abilities". Speed factors have been readily identified in factorial studies.

In 1963, French and his co-workers assembled a kit of reference tests, containing 24 primary factors which had emerged in over 120 studies of ability. These factors have been repeatedly observed and thus have the status of well-replicated dimensions. Speed factors of various types were identified. These are listed below -

"Established"Primary Factors (French et al. 1963)

Speed of Closure:	Ability to integrate disarranged visual patterns.
Spatial Planning:	Ability to quickly survey a complex spatial field and for example, trace a pathway through it,(e.g Maze Tracing Speed).
Perceptual Speed:	Speed in identifying pattern in visual material.
Number Facility:	Speed and accuracy in numerical tasks.

The French et al. (1963) list, according to Horn (1972) is somewhat conservative. For example, it does not include "Speed of Reaction" which has emerged in a number of studies (Horn 1972), nor does it include as a potential source of factors those which could be detected in individually administered tests. All the factors listed by French et al. (1963) are derived from group tests.

There have been many large scale studies that have isolated speed factors of one type or another. Kelley (1928) studied the inter-correlations of batteries of tests given to groups of pupils at different ages and found speed factors at each age. Thurstone's studies of the primary mental abilities (Thurstone 1938;Thurstone and Thurstone 1941; T.G.Thurstone 1941) at different ages have yielded factors such as 'P - Perceptual Speed' for some but not all age groups. T.G.Thurstone (1941) reported that in the study of 710 8th grade children, ten factors were extracted by means of the centroid procedures and after rotation, only six factors were clearly identified. Factor P, found in studies of both College students and high school children did not emerge in the data for the young children. However, the N (Number) factor, described as "the ability to do numerical calculations rapidly and accurately" was found in the youngest group.

The United States Employment Service has produced a series of aptitude tests (General Aptitude Test Battery) based on extensive studies of large numbers of subjects who were applying for jobs of different types. Among their stable and consistent factors are included Motor,Perceptual and Clerical Speed (Vernon 1961). Perceptual and Psychomotor Speed have also been identified in the various studies carried out by Guilford and his colleagues using Air Force trainees as subjects (e.g Guilford 1940). A speed factor has also been found in the Spearman-Holginger Unitary Traits Studies. Holzinger and Swineford (1939) for example repeated the earlier investigations which had identified a group factor of speed in four tests (additions,coding,counting dots, and distinguishing between straight and curved capital letters). They tested 7th and 8th grade children of both sexes in two schools (301 subjects) using group tests. The data for the two schools were analysed using the bifactor method and the solutions were examined for sex differences. Their speed tests "designed to measure mental or perceptual speed" revealed some significant sex differences, the girls being quicker than the boys.

The investigation carried out by Davidson and Carroll (1945) was an attempt to establish the linear independence of speed and level scores and to examine their relationship with time-limit measures. Level was taken to

TABLE 5. Factors from Davidson and Carroll (1945)

Factor	Nature
A	Speed of computation
B	Level of reasoning
C	Speed of reasoning
D	Uninterpretable
E	"
F	General speed factor

mean the amount the subject knows and speed the rate at which he works. Speed scores were the times taken by subjects to attempt every test item once. Level scores were the number of correct answers when subjects were allowed to take as much time as they wanted. Time limit scores were the number of correct answers in a prescribed time.

These authors used a variety of tests including the Revised Alpha Examination (Form 5) and the Minnesota Speed of Reading Test for College Students. These tests were administered to a group of 91 subjects (12 men, 79 women), all psychology undergraduates. Timing was carried out using a large clock with a sweep-second hand placed in front of the group. Time limit testing was followed by the students being instructed to work rapidly to the end of the test. They were then required to note the time taken on their answer sheets. The subjects were also allowed to correct their previous answers using a red pencil.

The skewed time scores were transformed to near normality by a reciprocal transformation, and the data were factor analysed by the centroid method and then rotated to simple structure. Six factors were extracted (see Table 5) and in addition the relationship between speed, level and time limits was studied by examining certain multiple correlations.

On the basis of the multiple correlations between Speed-Level and Time Limit measures, it was found that for some tests, speed made little contribution, for others that level was unimportant and for a third set both made equivalent contributions, indicating the factorial complexity of time-limit scores.

The unrotated loadings of the second factor ~~is~~ are particularly interesting in that ~~it~~ they appear to define a bi-polar factor which differentiates speed and level measures.

Myers (1952) applied different forms of a non-verbal reasoning test to 600 recruits in their first term at College. Each of the forms was differently speeded by varying the number of problems to be done in each part of the test, 12 minutes being allowed for each part. A total of 19 different scores was obtained for each form of test, including number correct, number wrong, number skipped and number attempted, the latter being measured by the last item attempted. (Lord (1956) found that this last measure was particularly prone to 'dishonesty' in the testees).

Each of the three forms of the test was analysed separately and each produced two orthogonal factors which Myers described as the tendency to answer correctly and the tendency to respond quickly. The former factor was found to be somewhat more valid for predicting course grades.

Howie (1956) used 13 tests (from which he generated an additional 9 measures) in his study of speed and accuracy in 158 12-year old boys. He does not describe his procedures in detail. The purpose of the study was to examine the relationship between speed and accuracy when ability was held constant. A centroid analysis was used and three different orthogonal rotations undertaken, mainly because the first two could not resolve certain anomalies in the solution.

'v:ed'

The first factor appeared to be a mixture of 'g' and ' which could not be partitioned. The second appeared as a general speed factor not restricted to simple or routine tasks. All the accuracy scores loaded negatively on this factor and no other accuracy factor appeared. Howie states that for his data, with ability held constant, "faster workers will make more errors, slower workers will make fewer errors". The third factor appeared to be one of verbal ability and the fourth was uninterpretable.

Lord (1956) studied a group of 649 entrants to an American naval college. The ages of this group were not reported and it has to be presumed that they were highly selected. These subjects were given seven tests of verbal, spatial and arithmetic ability. The entire battery, together with other data, contained a mixture of tests varying in their degrees of 'speededness' as well as 'level' tests "involving virtually no speed". Also included in the battery were six reference tests. All test data, as well as information on school grades (a total of 36 variables) were factor-analysed using Lawley's Maximum Likelihood procedure.

In addition to the standard scoring procedures, a 'last-item-attempted' score for one of the speeded tests in each area was computed. Lord found that subjects who completed the speeded tests in fact skipped items or else answered at random. Although the per cent of examinees who finished was low (2% to 11%), he does not state how these data were treated in the computation of scores.

After extracting 10 factors (of which 9 could be interpreted), further analysis was abandoned because of the statistical non-significance of the residuals. The factors were then rotated "with the help of a matrix rotator" to produce meaningful oblique axes. In addition to the anticipated verbal, spatial and mathematical reasoning factors, four speed factors were found (verbal, number, perceptual and spatial speed). These factors were intercorrelated "demonstrating the existence of a general speed factor at second-order level". With one exception, which Lord regarded as minor, all correlations between course grades and the four speed factors were positive but low.

Lord's study is in many ways typical of factorial research which gives little attention to the quality of data subjected to factor analysis. Attention has already been drawn to the limitations of four of his variables and other aspects of his study are questionable. For example, he asserts that the course grades were "virtually unspeeded", his reason being that "almost every student finished". Such statements ignore the speed stress which is inherent in the examination situations that are used to generate the data.

Porebski (1954,1960), in a series of investigations, challenged Spearman's (1927) view that it is "unnecessary to distinguish between speed and quality of thinking on the assumption that these two characteristics correlate almost perfectly". In his view, Spearman (1927) used an inappropriate procedure when simple tests were scored first on speed and secondly on accuracy without time limit. As Porebski (1954) put it:"It can be doubted whether an accuracy score on a simple test is adequate to measure the quality of thinking". To test his views that speed and power are distinct factors of ability and that they are fundamental (in the sense that they over-ride abilities such as verbal, spatial and numerical), Porebski constructed six speed and three power tests. Each test series was sub-divided into verbal, spatial and numerical tasks. The speed tests had to be completed within a specified time in the test room whereas the power tests were taken home and the subjects allowed up to two weeks to solve the problems. Whereas the speed tests allowed for a wide range of scores, the power tests were graded into three levels of solution, correct, partially correct and wrong. These tests were given to 50 subjects, the scores correlated and factor analysed. A centroid method was used and the loadings rotated. The orthogonal solution produced factors which were described as 'g' speed and 'G' power and specific factors of speed.

Porebski's 1954 paper, and a subsequent study (Porebski 1960) have been subjected to detailed critical analyses by Vincent (1955;1960) who put forward a number of reasons for disregarding Porebski's conclusions. The limitations he identified include inappropriately chosen tests and the fact that, with only nine tests given to such a small group,"it would not be possible to extract more than two factors". Vincent(1960) also advanced the view that the 'power' factor could have arisen from a variety of other determinants, such as persistence. Porebski has made no systematic attempt to refute this or some of the other criticisms.

Finally, the method of administration of the power tests, while certainly not subject to the criticism of being time-limited, nevertheless does raise other questions which add to the limitations of this study. For instance, we are not informed how Porebski insured that the power tests were solved by the same subjects who solved the speed tests. While Porebski has repeated his tests a number of times (Porebski 1954,1960) the same limitations would seem to hold for these repetitions.

Mangan's (1959) factorial study was an attempt to answer a number of questions on the relationship between speed, power and various temperamental attributes.

Specifically, he was concerned to see if it was possible to identify speed, level, persistence, accuracy and carefulness factors under what he termed "ordinary experimental conditions". He meant by the latter, group tests to groups of 30 or so school children (sex unspecified) aged 13 to 14 years in a London Comprehensive School. In addition, Mangan sought to discover whether speed or level was the more appropriate measure of the general factor, and to determine the relative contributions of speed, persistence, level, accuracy and carefulness to the power test variance.

A total of 200 subjects was tested over a period of eight weeks, with test sessions lasting up to 60 minutes. The battery consisted of 32 tests which yielded 38 measures. Thurstone's centroid method was used to factorize the battery and the factors rotated to produce an orthogonal solution with the sole constraint being the retention of as large a 'g' factor as was possible.

Included in the large battery of tests were scales aimed at measuring "clerical,perceptual and motor speed and fluency", as well as tests of verbal and motor tempo. Tempo was defined as "the natural rate or rhythm, or basal speed". Although Mangan concedes that timing of individual items on the speed and tempo tests "was considered desirable", he used only gross rate measures. It is also worth noting that on 14 of the tests, the subjects were required to record their own times to completion, a procedure not conducive to accurate timing.

Seven factors accounting for 40% of the variance were extracted. These are listed below -

'g' :	a general power factor
Persistence:	mainly on power tests
Perceptual speed:	perceptual,verbal and non-verbal and clerical speed on easy material.
Speed-accuracy:	bipolar factor contrasting motor speed,writing and tempo,with accuracy.
Number:	mental and problem arithmetic
Verbal tempo:	difficult to identify
Fluency:	of doubtful significance

In discussing his results, Mangan notes that he had some difficulty in deciding whether or not the general factor was 'g' or a 'g+v ed' combination. Whereas the untimed power tests had higher loadings than

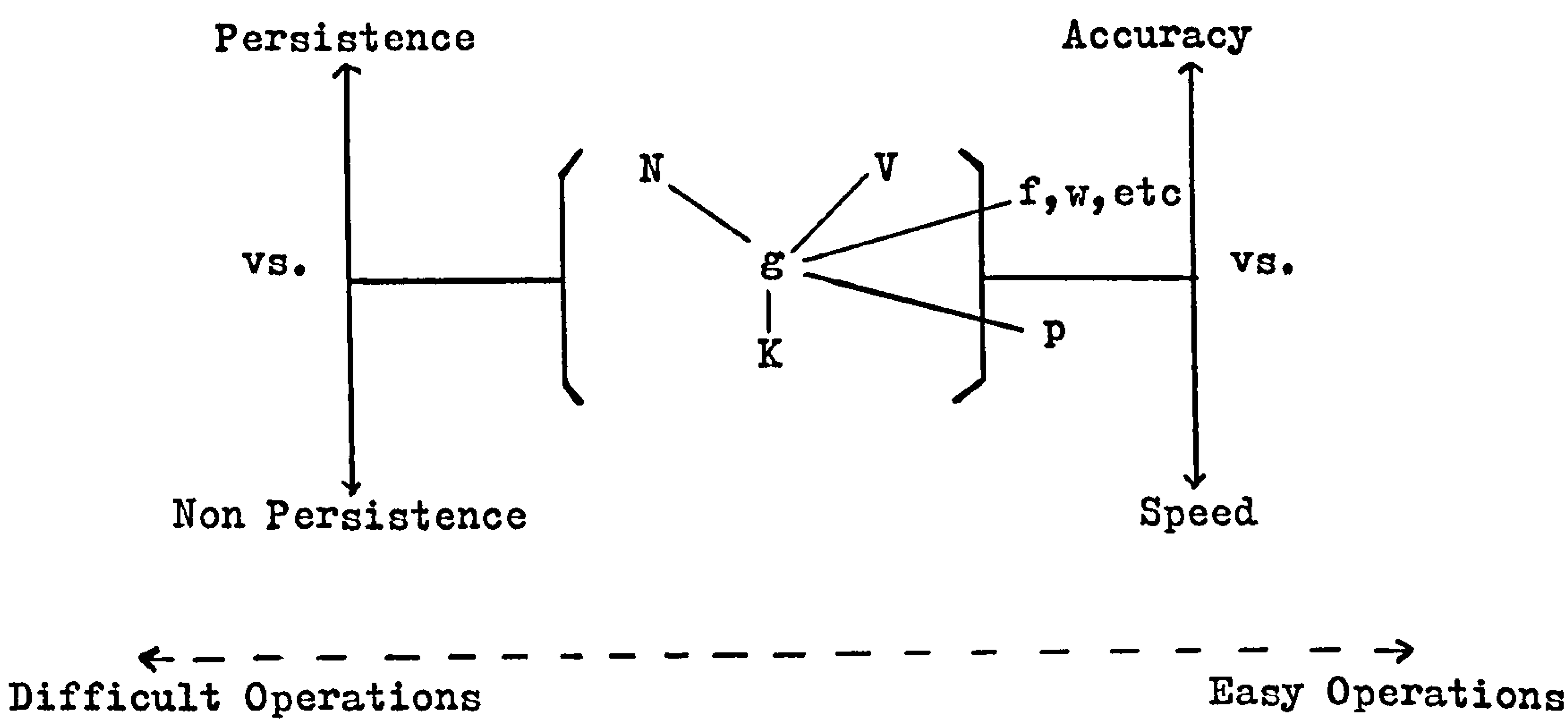


FIGURE 5. Mangan's (1959) schematisation of his findings.

the timed power tests on this factor, they were also the more lengthy. Two of the timed power tests did nevertheless have higher loadings (.55,.48) than other timed tests. Speed, as measured by Mangan appeared to be of "minor importance at a difficult cognitive level", although here it must be recalled that 'speed' was defined by rate measures. Mangan had schematized his findings as shown in Fig.5.

According to his view, the content factors cannot be measured in isolation from what he has called the two work attitude factors of speed-accuracy and persistence. Thus, for difficult material persistence enters into performance but is absent particularly when the tests are highly speeded. Speed-accuracy on the other hand enters into performance only when easy or unskilled tasks ate presented. If power tests are given with unlimited time, the outcome, according to Mangan, will depend on the temperamental attribute of persistence. If highly speeded tests are used, speed-accuracy will intervene. When there is no time limit , the work-attitude of the subject will change from speed-accuracy to persistence as the test material moves from easy to difficult tasks. Mangan also emphasised that the other speed factors were relatively distinct from speed-accuracy.

Mangan's conclusions, while seemingly sensible, are too firmly presented, given the narrow age range of his subjects, the fact that possible sex differences were not taken into account and given the crudity and possible inaccuracy in some of his measures. There is, for example, some evidence that factor structure will vary as a function of the sex of subjects (Cooley and Lohnes 1968) in a group not much older than those tested by Mangan, in a similar setting and when tested in groups.

Clarke and King (1960) administered a battery of three motor and three perceptual speed tests to a group of subjects who were available to a Highway Traffic Safety Committee. 106 of these subjects were being re-examined because of "excessive violations", the other 93 being routine license applicants. Subjects ranged in age from 18 to 77 years and were all males. The group was also heterogenous with respect to years of education (3 to 21 years).

The WAIS vocabulary test was also given to these subjects. The intercorrelations were factor analyzed using the principal axes method. The three main factors were rotated to orthogonal simple structure but the third factor was not interpretable following rotation.

The rotated factors were designated perceptual and motor speed. Using age,education and vocabulary as reference variables, Clarke and King suggest that slowness in selecting an appropriate response, the perceptual component, was inversely related to "verbal ability" while slowness in making a response, the motor component was positively associated with increasing age.

The same analysis procedure was used with education partialled out of the correlation matrix to ensure that the observed structure was not attributable to the educational heterogeneity of the group. Only minor alterations in the loadings were observed.

Chown's (1961) factor ahalytic study of "the rigidities" and their relationship with age has yielded data on the 'speed factor'.She used 16 tests of rigidity and two measures of intelligence (20 minute Matrices and the Mill Hill) and the battery was administered to 200 men ranging in age from 20 to 82 years. Principal component analyses on the data for the full groups were followed by similar analyses on the data for theee sub-groups (Young, middle-aged and old men). Chown carried out graphical rotations to remove two factors, those of non-verbal intelligence and age, from all the factor analyses. Hotelling's principal component method was used.

The measures of speed in Chown's study were derived from a number of the tests and varied somewhat in their nature. For example, speed of forward writing was one index of speed, counting letters (é.g. fourth letter after C) another. She found, on examining the inter-relationships among the factors, that the most marked changes with age were those associated with speed. As she states "the speed tests formed their own unique factor in the young group, maintained this to a lesser extent in the middle group, and were loaded most highly on the Non-Verbal Intelligence factor in the old group. Thus among old people, but not among the young, these speed-tests became a measure of intellectual capability and of the extent of the preservation of this function". The extent to which these findings are a function of the timed administration of the Matrices is not known. Chown was investigating this possibility when her paper was published.

In the late 1950's,Flanagan and his colleagues (Flanagan et al.1964) initiated a massive study of adolescent 'talent'.Its purpose was to gather material on abilities and motives to be used in career guidance. 'Project Talent' as it was called, was based on a multifactor conception of human talent,'general intelligence' being regarded as too oversimplified a concept for such a purpose. A large battery, comprised of 60 different ability and 38 motive scales was devised and administered over two days to about 440,000 students in the 9th to 12th grades. The group testing was carried out in 1,353 public,private and parochial schools in all parts of the U.S.A. A stratified random sampling procedure was used to select schools and the sample was made up of about 5% of high schools in the U.S.A. Since the publication of the initial report of the project (Flanagan et al.1964),further analyses of the data have appeared (Lohnes 1966;Cooley and Lohnes 1968) and there are plans to follow the subjects into the 1980's.

Some of the tests given are presented in Table 6, together with

TABLE 6. Project Talent tests.

Time*	Test	Content
90	Information	General information
9	Arithmetic computation	Speed and accuracy of basic skills
11	Abstract reasoning	Non-verbal reasoning
3	Clerical checking	Speed, accuracy. Verbal content
3	Table inspection	Speed, accuracy. Numerical material
3	Object inspection	Speed and accuracy in detecting differences in objects
3	Preferences**	Speed of decision making
52	English	Test of English usage
4	Visualization(2-D)	Spatial orientation of two dimensional shapes
3	Disguised words	Deciphering of codes

* Minutes

** The Preferences Test is an experimental version of a test requiring the testee to select which of a pair of adjectives best describes a friend.

TABLE 7. Project Talent Ability Factors (Lohnes 1966)

Factor Name	Percent Variance
Verbal Knowledges (sic)	18.7
Grade	7.8
English Language	6.6
Sex	5.7
Visual Reasoning	5.3
Mathematics	4.1
Perceptual Speed and Accuracy	3.6
Screening*	3.3
Hunting/Fishing*	2.2
Memory	2.1
Color/foods*	1.9
Etiquette*	1.6
Games*	1.5

*These factors are components of the test which sampled a range of areas of specific knowledge on the Information Test.

the time limits employed for each and a brief description of the test materials.

A few of these tests were designed so that nobody would finish and another few were less 'speeded'. Most of the tests were designed to have as generous time limits as were possible in the two-day test sessions. All tests were group administered.

Although the data were gathered for all subjects, most of the analyses were carried out on specially selected representative sub-groups. These subgroups did however contain substantial numbers of subjects (7,000 to 16,000 in some analyses).

Lohnes (1966) has described the outcomes of a series of principal component analyses of the data. The varimax rotation prodedure was used to produce orthogonal solutions. This approach was adopted so as to accord with the basic multifactor theoretical conception of the study. The original 60 ability scores were reduced to 13 uncorrelated factors accounting for 64.6% of the variance. As can be seen from Table 7 two of the factors (sex and school grade) which emerged were not part of the test battery.

The outcome of the factorial study on 17,000 subjects is presented in Table 7, together with the percentage of variance accounted for by each factor.

Three of the factors, Visual Reasoning, Perceptual Speed and Accuracy and Memory, were considered to be "differential aptitude factors". Lohnes (1966) defines an aptitude in terms of a performance set that facilitates speed and precision of responses,"to a specific, unique class of relatively simple tasks" (pp.4-1 to 4-2).

Perceptual Speed and Accuracy (PSA) was loaded by four "highly speeded tests" (Clerical Checking, Table Inspection,Object Inspection and Arithmetic Computation). While the loadings of these tests were somewhat low, Lohnes (1966) attributed this to low reliabilities "brought about by widespread discrepancies in the timing of the tests in different schools" (p4-9). Apart from Arithmetic Computation, considered by Lohnes (1966) to be a complex test, all these tests showed low intercorrelations with the other tests in the battery.

Separate analysis of the data were undertaken for the two sexes and the different grades. In the analyses contrasting males and females PSA emerged with a different structure in the two groups, with the sex differences being most extreme in the 12th grade. Males overall were found to be superior on the Preferences Test whereas females excelled on Arithmetical Computation, Table Reading, Clerical Inspection and Object Inspection.

Two studies quoted by Cooley and Lohnes (1968) also reveal some important findings on PSA. Both these studies were carried out on Project Talent testees but were issued as internal reports to the project. In the first of these, Shaycroft retested 7,500 subjects in the 12th grade who were first seen as 9th grade students. It was found that the basic ability structure remained stable. However, PSA emerged with "very low" stability coefficients, .33 for males and .41 for females. Cooley and Lohnes (1968), in commenting on these coefficients assert that these findings are attributable to "anomalies in retest administration". It should be noted that a similar reason was given to account for the low reliabilities when these tests were first administered (Lohnes 1966).

The second study, carried out by Schoenfeldt, consisted of an analysis of the heritability of the various TALENT factors. The original sample contained over 2,000 same sex twins, By means of a specially developed questionnaire, Schonfeldt isolated 150 male and 187 female monozygotic pairs and 53 male and 103 female dizygotic pairs; sex differences in heritability were found in some of the factors. Of particular importance in the present context was the failure to find a statistically significant heritability coefficient for the male students whereas an h^2 of 0.59 was found for the females.

d. Conclusions from factor-analytic studies.

The problems which have served to focus factor analytic studies have been sufficiently explicit. They include the questions of the presence of a speed factor at general, or group level, the relationship between such factors and factors of power or 'g', whether or not such factors are a function of difficulty level of the test material, their relationship with accuracy, persistence and so on.

Any attempt to draw together the findings of factor analytic studies of speed is hampered by a number of methodological shortcomings. Basic to the difficulties is the quality of the data, a point which has already been considered in some detail. Further complications are introduced by the specific procedures used to analyse the data, which is in turn related to theoretical pre-conceptions and the type of rotation employed. At a higher level are found the problems of factor interpretation which further confound the task of making sense of the studies. Also, as has emerged in the preceding review, the strength and patterning of the relationships appears to be a function of age and sex. There have been no systematic investigations of the role of intelligence, education and social class of subjects. Finally, what is meant by 'speed' in some factorial studies does not coincide with other definitions of speed.

Taking an overview, it is also somewhat disconcerting to find an absence of systematic replications. The overriding impression is that most studies are of the "one-off" type, Porebski (1954;1960) being possibly the only exception.

Vernon's (1961) conclusion seems to be the most appropriate at this point. After reviewing the various studies of speed in the factor-analytic literature, he reaches the inevitable conclusion that only when further research has been conducted will it be possible to make a proper assessment of speed and its relationship with other human abilities. There does however appear to be sufficient evidence that a variety of tests show low but consistent inter-correlations. These are predominantly of the repetitive item type and are of low conventional difficulty, termed cognitive speed tests by Cattell (1971). There is also some evidence of sex differences on these tests in terms of 'level' and possibly in terms of pattern as well. How this speed factor relates to intelligence within sexes and between ages at varying levels of ability is at best uncertain.

7. Studies of Speed Using Item Times:

The studies most pertinent to this thesis are those in which individual items have been timed. While such studies can be divided into those using group - as opposed to individual - test procedures, both types will be considered in this section.

In 1925, Hunsicker conducted what was probably the first study which used individual item times in investigating 'rate and ability'. In fact she employed a variety of procedures to gather her data, including group-testing . However, because of the problems associated with group tests, Hunsicker discarded these data and used only what she called "individual-test" data. These data were actually obtained from groups of 4 to 7 subjects tested together. Hunsicker timed both individual items and blocks of items. In all, 368 subjects were tested on items of varying difficulty. To control for accuracy, only correct items were used. Subjects were of both sexes and ranged from school children to college students. All subjects were given arithmetic items and sentence completion tasks.

On averaging the correlations between rate and level across age groups, the values ranged between 0.39 to 0.61. Hunsicker concluded

> "These values leave ho escape from the conclusion that for the groups studied under the conditions of this experiment there is a consistent and fair positive relationship between rate and ability".

Hunsicker suggested that had there been a wider range of ability in her groups, the correlations would have been more substantial. Correlations between rate measures ranged from 0.57 to 0.81, and those for level measures between 0.55 to 0.71.

It is difficult to determine the accuracy of Hunsicker's timing procedures and what effects the act of timing would have had on the subjects. It is likely that the test situation produced conditions of stressed testing.

The study by Peak and Boring (1926) is in some ways a classic investigation of mental speed. Their report presents an important theoretical analysis of mental speed (discussed in a previous section), and their study was among the first to use individual item times. It's major limitation was the use of only five highly selected subjects, (2 males,3 females), who were either advanced undergraduates or psotgraduates in psychology.

All subjects were given Forms 5 and 6 of the Army Alpha and Forms A and B of Otis. Each subject was tested individually and the tests administered in the same order. The normal time limit instruction was omitted but subjects were instructed to work quickly and accurately and to indicate abandonments. Each item was timed by the examiner, apart from one sub-test where 5 consecutive solutions were timed.

To investigate individual differences in speed, only correct solutions were examined. Even this index, as the authors note, is not quite adequate since "the subjective assurance of the subjects must have varied in the different items". On the basis of the data used in this analysis (which is too extensive to present here in detail), Peak and Boring observed that solution times showed significant individual differences that appeared to be fairly constant.They concluded that speed differences inhere in the single item, rather than between items. As they state "the tendency for differentiation as to speed appears in the simple intelligent act represented by a single item...." There is however a difficulty with this conclusion in that they do not clearly state how items were timed. For example, if they simply noted the time at which the solution was recorded, then individual item times will include interstitial times. If the timer was reset for every item and timing initiated only when the subject initiated the problem solving process, the interstitial times would not be included.

Peak and Boring also gave their subjects a visual-reaction time task (simple) and found that the average time for 100 reactions correlated .70 with a weighted average item time on the Otis and .90 on the Alpha tests. Further, the scores obtained by the subjects in the standard time allowed on these tests both correlated .90 with the reaction times.

It should be noted that Peak and Boring were cautious in discussing the implications of their findings, firstly because of the small number of subjects and also because they only examined the central teddency and not the variability of their data.

Sutherland's (1934) report presented a series of studies, the details of which are too extensive to describe here. His measures of speed varied according to the task given. For example, on a series of performance tasks, each test was timed, but because Sutherland presumed that such times would be too unreliable, each time was converted to a deviation score and the median of these was used as the index of speed. For his university student samples on these tests, the Spearman-Brown reliabilities were much too low to allow serious consideration of the data. The reliability coefficients for speed on the Blocks, Formboards and Cubes were .68,.41 and .62 respectively.

In another test series, Sutherland used a timing procedure whereby 3 large cards with single digit numbers on them were turned every two seconds. At certain points in the test, subjects were asked to note the numbers on the cards currently displayed. Thus, he could only obtain fairly crude rate measures for tasks such as cancellation of letters, adding groups of 3 digits, simple arithmetical problems, and the like.

On the basis of his data, Sutherland could find no evidence of a speed factor that is general. It is only when problems are of low difficulty that "a factor of speed comes into operation." Sutherland goes on to conclude

> "If intelligence is taken as the general factor 'g', then no specific factor of speed is involved in solving problems whether these are easy or difficult".

Slater (1938), in his study of speed, used an adaptation of Thorndike's CAVD as well as a number of other tests. His subjects were asked to work at their own rates. About 150 subjects aged approximately 14.5 years were tested in groups. Equal numbers of males and females were used. A cyclometer (described in a previous chapter) was used to indicate the time. Subjects were told the purpose of the machine and were asked to put down the time at which they began each question. Further, they were asked to raise their hands when a new set of questions was required. Thus, despite the intention of the author, the emphasis on timing and the visible progress of others makes it highly unlikely that Slater was getting measures of subjects working at their own rates. If anything, this study appears similar to Furneaux's (1955) stressed speed condition.

Rate of work measures were computed from the correct solution times to control for accuracy. Measures of level were obtained from the number of correct solutions on the level test. Slater found that although the speed measures correlated among themselves, they did not show any close association with the various measures of intelligence, irrespective of whether or not these tests were timed. Correlations were found to vary between 0.0 and 0.4

depending on which subjects were considered. He also found that subjects tended to have consistent work rates irrespective of the type and difficulty of the tasks. These rates, according to Slater "could not be considered to depend only on the amount of their general intelligence".

The procedures of this study make it very unlikely that the time measures were accurate. Slater himself recognised this shortcoming. Slater's procedure for measuring average speed on each of the tests resulted in a degree of data loss that, because of its differential nature, further complicates the interpretation of his data. For example, the average rate of work measure was computed for only those series in which the subject had "successfully answered any considerable number of problems..." Slater did not specify what he meant by 'considerable'.

In a carefully devised study, Tate (1948) examined differences in speed of response to items at different levels of difficulty. His immediate concern was the extent to which speed of response to mental test items differentiated individuals once the effects of accuracy were controlled. He also attempted to discover a factor of speed independent of altitude and the function being measured.

Thirty -six high school students aged 188mos.on average (s.d 7.4mos.) were used as subjects. This group, as judged by their mean mental age, were of above average ability (mean M.A 218.7mos.,s.d.16.7mos.) on the S-Form of the California Test of Mental Maturity and all were enrolled in classes to prepare them for college entrance. Although the group was made up of both sexes (14 females,22 males) only minor sex differences appeared and of these only one was statistically significant. All subjects were randomly selected but the group did not consitute a representative sample of high school children. They were all students in Tate's class who were willing to spend three to five hours doing tests.

Tate used four tests,Arithmetic Reasoning,Number Series,Sentence Completion and Spatial Relations. Each was divided into three non-overlapping levels of difficulty and the battery administered during the vacation and at weekends. Subjects were instructed to complete items as quickly and as accurately as they could, but Tate does not give details of the wording. All items were individually timed.(The procedures have been described in an earlier chapter). Times were recorded in seconds and transformed into logs to satisfy the distributional assumptions for the analysis of variance used on the data.

Tate found that when difficulty and accuracy were controlled, there were highly significant individual differences in speed. Controlling for accuracy differences, it was found that those subjects who were fast at one level of difficulty were also fast at other levels. Also, subjects who were fast in one test content area were also fast in the other three. For difficult items, using an independent measure of altitude in the same content area, correlations between speed and altitude did not differ significantly from zero,indicating that mental speed is independent of altitude. However, in addition to this general speed component, the items "were eliciting a special ability in speed linked to the function in which it was being measured". Finally, Tate concludes "and that there was an independent factor of speed operative throughout a wide range of difficulty appeared statistically certain".

Although there were some serious methodological limitations, there is little doubt that Tate's study was a major advance in the investigation of mental speed. The restrictions attached to his findings are mainly those of generalizability. The narrow age and ability ranges and the above average ability of the subjects are some of the features which preclude any wide generalization of Tate's conclusions. Other limitations arise due to his timing procedure and the failure to counter-balance the order of testing. Some further restrictions must also follow from the fact that the person who conducted the study was also the class teacher. It would also be important to discover if instructions which stressed neither speed nor accuracy and a surreptitious timing procedure would lead to the same conclusions.

The study published by Cane and Horn (1951) was indirectly concerned with speed in relation to ability. However, it employed an interesting procedure for recording solution times and was well conceived in terms of its design. Cane and Horn were mainly concerned with developing a set of parallel tests matched for type of question, difficulty and average solution speed. This provided them with an opportunity to investigate the relationship between speed and total score, and between difficulty and time spent in doing any item.

Cane and Horn used the shapes Analysis Test made up of six types of question about 2- and 3- dimensional shapes. The subjects were 13 -and 14-year old children at a Secondary Modern School. There were 60 children in each group, with equal numbers of boys and girls. The timing procedure permitted the recording of time from item presentation to the time at which the subject initiated an action which led to the presentation of the next item. Details of the apparatus were described in an earlier chapter.

The test session lasted an entire day, with three breaks between periods of testing. Each subject answered 108 problems and also completed the A.H.4. The authors were careful to design the order of testing (balanced incomplete blocks) to control for order and other effects. Time scores were converted to logs.

It emerged that the tests given were much too difficult for these groups of subjects. The scores (number correct) were generally less than one half the maximum score, with many scores of zero and very few near the test ceiling.

Cane and Horn report that the position of a set of problems did not affect the score. Males were significantly better than the females on the total score measure and there was also a significant sex x age interaction.

	BOYS	GIRLS
13 years	1040	786
14 years	1095	979

Although the authors describe this interaction as showing that girls improve more than boys, it is not a true 'improvement' because different age groups were used in this 'cross-sectional' design. In any case, given the difficulty of the tests, it is equally likely that the age differences in scores interacting with sex of subject is artefactual as it seems likely that both groups of boys were at their score celing, whereas the data suggest that the scores of girls could still increase on these tests.

Cane and Horn also found that the time spent on a set of questions was significantly affected by its position. There was a general practice effect on speed, and although boys were slower, there were no age differences. The effect of practice was to considerably diminish the time spent on a test. However, there was no effect on the score obtained. This may well have been a consequence of the difficulty of the tests. Also, although the better subjects (higher total correct) were slower, the relationship observed was not linear, as it was found that the very slowest subjects also had the poorest scores.

Unfortunately, the various speed scores in this study were not correlated with the scores on the A.H.4. The Shapes Test showed only a low correlation (.29) with the A.H.4 but the many zero scores and the few scores in the upper score range make it highly likely that this correlation is artificially low.

Cane and Horn present an interesting hypothesis on the basis of their data. They suggest that when a difficult test is given with a time limit, poorer subjects will cover more of the test and so approach nearer to their maximum score than do better subjects, the latter being slower. If the test is readministered, all will tend to work faster. But, according to their observations, "the increase in the number of questions answered will be of greater benefit to the good subjects, who will get more of them right,so the good subjects will improve their scores more than do the worse ones".

Apart from the present investigation, the study by Brierley (1969) is the only attempt known to this writer to test Furneaux's approach to problem solving. Brierley's study is distinguished by its careful attention to methodological and procedural problems. Unfortunately, its length (over 300 pages excluding tables and appendices) precludes any description of the full study. However, as is noted in other sections of this report, he encountered a number of difficulties with Furneaux's item scaling procedures. This led to some limitations being attached to his findings and these should be borne in mind when considering his results.

There were three main components to Brierley's study. The first was the procedure used to record item solution times. His automated testing procedure has already been considered. The second was his attempt to employ Furneaux's item scaling procedures. This is discussed in greater detail in a later section. The third component, concerned with his findings on speed and accuracy, is presented here.

In brief, Brierley found that Furneaux's theory of intellectual speed measurement held up to some extent when investigated in normal subjects by means of an automated individual test procedure. Clinically diagnosed neurotic subjects displayed problem solving characteristics that were substantially the same as those of normal subjects. The normal subjects in this study were hospital personnel (administrative staff and nurses aged 15 to 45 years) who were predominantly of above average ability. No attempt was made to examine the data for normals for differential patterns of performance with regard to the sex of the subjects.

Brierley also proposed a dimensional system for 'power' measurement based on Thurstone's (1938) views. On the basis of his power scores, he was able to show that neurotic subjects were significantly less efficient than normal subjects. Further, extraverted neurotics tended to be quicker than introverted neurotics and that for a considerable range of item difficulties they were less accurate.

Chown and Davis (1969) selected items from the Nufferno Letter Series (level) for their study of the effects of age on speed and level of performance. 10 subjects aged 45 to 70 years and 12 subjects aged 20 to 35 years were given the items. The proportion in each passing an item was used to establish item difficulty for a sub-set of the items. From these data, they selected 10 items representing five difficulty levels and these were then arranged as two cycles. This modified test was then administered without time limit to 42 subjects, half of whom were between 20 and 49 years of age, and the other half, between 50 and 69 years. The groups were matched on Mill Hill Vocabulary scores and the tests were individually administered and timed. Further details of timing and other procedures were not given.

Chown and Davis (1969) found that the two age groups did not differ in mean number of items correct. No information on the ratio of errors to abandoments is presented.

Solution times were transformed to log values for the correct solutions. Older subjects were found to spend more time on easy items, rather than on those that were more difficult. Older subjects also showed greater persistence on items which they solved incorrectly.

A study such as this is difficult to interpret for a number of reasons, lack of information on timing procedure, and the difficulty scaling being two of the major reasons for this.

Yates (1963,1966) has carried out a series of investigations to examine the effects of time limits in test scores. In the 1963 study, using the Standard Matrices, he found a group of students who appeared to be slow but accurate. This group attempted fewer items than other subjects, but the majority of those attempted were correct. The 1966 study aimed to repeat this finding. A group of 86 first year engineering students were given 30 minutes to complete the test and were then allowed a further 15 minutes, not having been told beforehand that extra time would be allowed. The Nufferno Level Test GL/26 was then given to the group with the 25 minute time limit being made explicit.

Yates was again able to identify the slow but accurate group. Their level scores were the same as those of subjects who worked faster. Yates suggested that because his was a pre-selected group, the proportion of individuals in the general population with this approach to tests may be more extensive. While Yates also commented on the possible personality characteristics of such a group, no personality variables were used in the study. A similar performance pattern has also been reported by Hichens (1968), using Grammar and Intermediate school pupils. The sex ratios and ages of these groups were not described.

While the studies discussed in this section share the procedure of individual item timing, methodologically they are not much more adequate than the studies previously considered. The findings, taken overall, are inconsistent in some respects and consistent in others. Some authors have found that speed measures are related to each other, giving an indication of a more general speed factor. Tate (1948) also reports finding specific speed effects dependent only on the test. Others have found no general speed effect. Given that the criterion of speed varies between studies, that timing is surreptitious or overt, that instructions vary, and so on, it is perhaps not surprising that research produces inconsistent results.

8. Speed and Personality

Himmelweit (1946) studied a group of 100 neurotic patients diagnosed as either hysteric or dysthymic with a view to determining if they responded differentially on speed and accuracy tests. There were equal numbers of men and women in each diagnostic group and ranged in age from 18 to 40 years.

All subjects were given four paper-and-pencil tests and one motor manipulation task. For one half of each task they were instructed to respond both quickly and accurately and for the other half as quickly as possible. All subjects were also given Raven's Matrices and these scores were used as the basis for partialling out intelligence differences.

Himmelweit found a general speed factor as well as a general accuracy factor both of which accounted for substantial proportions of the test variance (ranging from 24 per cent to 41 per cent). On the paper-and-pencil tests there was no relationship between speed and accuracy but on the manipulation task a substantial negative correlation was found. Himmelweit attributed this difference between the two types of test to the presence of corrective feedback which was immediately available on the manipulation task.

Contrasting the performance of the hysterics and dysthymics, Himmelweit found that the former preferred 'speed', the latter accuracy. However, score differences between the two groups were only 'suggestive' in the case of speed and very significant in the case of accuracy. It was also found that irrespective of diagnostic groups women were quicker and more accurate than men.

Himmelweit was cautious about generalizing her results because of the clinical status of her subjects. However, given the age range of her group, the conclusions are even more restricted by her failure to partial out any age effects in the data.

Foulds and Caine (1958) administered the Matrices and Porteus Mazes to a large group of female patients (aged 20 to 59 years) admitted to a psychiatric hospital. They were all told that they could take as long as they wished to complete the tests but that they would be timed. The patients were later divided into personality groups and although significant differences in total time emerged, no attempt was made to partial out the effects of age. Details of the timing procedure were not reported and the impact of the clinical status of these patients on test performance was not examined.

Eysenck, in a study published in 1959, attempted to test the prediction that on a test which required subjects to work for an extended period of time extraverts would show a decrement in performance towards the end of the test. This prediction was based on the notion that a long test satisfied the conditions of a massed practice task which would generate reactive inhibition more readily in extraverts than in introverts. The test chosen for this study, the Morrisby Compound Series Test, was individually administered to two groups of subjects, 19 extraverts (E score greater than 30) and 28 introverts (E score less than 17). These subjects were selected from an initial group of 137 adult "male and female neurotics". Each item was timed separately.

The two groups did not differ significantly in total correct scores, or in the speed with which all items were completed. However, on dividing the test into the first 45 and last 15 items, introverts were found to be significantly slower on the latter, when correct solution speed was examined. Extraverts were also found to abandon items more quickly at the end of the test.

While these results appear to support Eysenck's predictions, the paper does not provide details of the ages of the subjects and the sex ratios within each group. Neither timing procedures nor instructions are described, so that the conclusions from this study are difficult to evaluate.

Lynn and Gordon (1961) obtained measures of E and N from a group of 60 university student volunteers (the first to volunteer) ranging in age from 18 to 23 years. These students were given both forms of the Mill Hill Vocabulary Scale and the odd-numbered items from the Matrices (Standard). They were asked to work primarily for accuracy but to be quick at the same time.

Compared with the general population, this group was found to be more introverted and neurotic than the 'general population' and their E and N scores correlated -.32. As a consequence, the main analysis of relationships between measures of intelligence and personality was based on partial correlations. Neither N nor E had any significant correlation with Matrices total score. However time taken to complete the Matrices was positively correlated with N and the test for curvelinearity of this relationship just reached significance.

Both E and N were found to correlate positively with the vocabulary score, the correlation with E only being statistically significant. None of the correlations was greater than 0.36 in absolute value. The sex composition of this group was not reported and presumably the data were not examined for sex differences. Also, the method of timing and the circumstances of testing were not reported. Although these factors detract somewhat from the value of this study, it is also likely that any correlations with intelligence would have been attenuated, given the narrow ability band of the subjects.

Eysenck (1967a) reports a study by Jensen (1964) which found a significant correlation of -0.46 between E.P.I. extraversion scores and time spent on the Matrices (Standard). These results indicate that extraverts tend to work more quickly and while they are reported to have made more errors, the trend did not reach statistical significance. The document reporting this study is not generally available and the research cannot therefore be properly evaluated.

Two studies by Farley (1966a;b), using the same group of subjects, have examined relationships between speed measures and personality variables.

The subjects in both studies were 30 male and female nurses, orderlies and other hospital domestic staff and University students, mean age 29.37 years (s.d. 11.14 years).

In the first study (Farley 1966a) subjects were required to copy simple geometric figures and letters on to a specially prepared sheet of paper. The total time to complete the task was recorded covertly, the subject being given "non-anxiety-evoking" instructions (not published). Time measures were converted into logs. Subjects also completed the MPI , (only Extraversion scores were used), Taylor's MAS and the Need-Achievement scale of the Edwards Personal Preference Schedule (Edwards 1959).

Farley (1966a) found the following product-moment correlations:

Log Time with Extraversion	-0.42	(p less than .02)
Log Time with MAS	0.11	(not significant)
Log Time with Need-Achievement	-0.20	(not significant)

Visual inspection of the scatter-plots revealed no curvelinear relationships.

Farley (1966a) concluded that extraversion influences free-response speed but that measures of anxiety and Need-Achievement showed no significant relationships. The value of this study is limited because of the failure to partial out age, to examine the data for sex differences and the small number of subjects. It is also somewhat surprising that no attempt was made to examine the relationship with N. These scores were available as is shown by the use of N in the Farley (1966b) study carried out on the same subjects.

In the second study, subjects were individually given the Nufferno Speed Test under stressed conditions.

Subjects were divided into 3 E groups on the basis of low, middle and high E scores. They were redivided into similar N groups for a separate analysis concerned with the effects of N on speed. Farley (1966b) checked his data to find if there were any interactions between E and N and failed to find any such effects in the speed scores.

Using an analysis of variance and Schaffe's test for individual comparisons, significant differences were found between extraverts and ambiverts, extraverts and introverts, and ambiverts and introverts vs extraverts. A similar analysis for the neuroticism group speed scores showed the mid-N group to be faster than the low N group and the low-and high - N groups combined. No other differences were significant. There were no significant differences in the number of correct solutions in the groups, and age, sex, and Mill Hill Vocabulary I.Q's did not show significant differences in the sub-groups.

Farley (1966b) concluded that extraverts are faster than introverts on such measures of speed. The N-sub-group differences were interpreted as supporting the "inverted U" hypothesis under stressed testing. As he points out, the groups were not extreme scorers on the two dimensions, so that the effects of the personality differences are somewhat restricted. Some of the methodological problems of this study were similar to those of Farley (1966a).

Ley et al. (1966) investigated two hypotheses, firstly, whether there would be a significant departure from linearity when the regression of intelligence test scores on anxiety is tested and secondly, whether there would be negative correlations between measures of extraversion and intelligence. Both hypotheses emerged from a number of reviews on the relationships between E,N,anxiety and intelligence.

The sample tested consisted of 144 volunteers (mean age 46 years,s.d.13 years) who had previously completed the Progressive Matrices and the Mill Hill Synonyms Test, Form II Senior. No information is given as to the time interval between the administration of these tests and the battery of measures given for purposes of this study. The later battery consisted of the MPI, the IPAT Anxiety Scale, the TMAS, and the 16 PF (Form C). The only contemporaneous measure of intelligence was derived from the 16 P.F.

Of the nine correlations between N or anxiety and intelligence, seven were significant and negative. Only one was significant and positive (Mill Hill and the Cattell Anxiety Scale) and one non-significant. The largest correlation was .36 (absolute value). Four of the coefficients were found to be non-linear but none showed the hypothesized inverted -U pattern when the data were plotted. They also found that all correlations with extraversion, when significant, were positive rather than negative. An attempt was also made to examine the relationships between N or anxiety at different levels of intelligence. In these analyses no 'inverted-U ' was found. Although these investigators were aware of the possible effects of age on their data, no attempt was made to partial out its effects. Sex differences were not commented on.

In an attempt to test a series of hypotheses derived from Eysenck's theory, Russell (1968) administered a battery of group and individual tests to 94 male and 92 female children aged between 10.5 years and 11 years 10 months. In addition to the JEPI and a latter series test, Russell obtained scores indexing speed and accuracy on the Gibson Spiral Maze, a letter cancellation task, mirror drawing,paper and pencil aiming and pursuit aiming. Russell's procedure for timing individual items on the letter series during group administration has already been described.

On the basis of his analysis, Russell reported that speed of performance was general across both motor and cognitive tasks. Accuracy was much less general. Speed and accuracy were negatively related for motor tasks but positively related on the letter series in males but not in females.

For the female subjects, speed of problem solving was significantly related to E, with extraverts being faster. However the correlation between speed of problem solving and E diminished as the length of the task increased. For males, E was unrelated to speed or accuracy. A sex difference was also found for some of the correlations with N. For females, the higher their N scores, the faster they tended to be.

Russell was unable to find any meaningful linear relationships between N and performance on the motor tasks for either males or females. For males

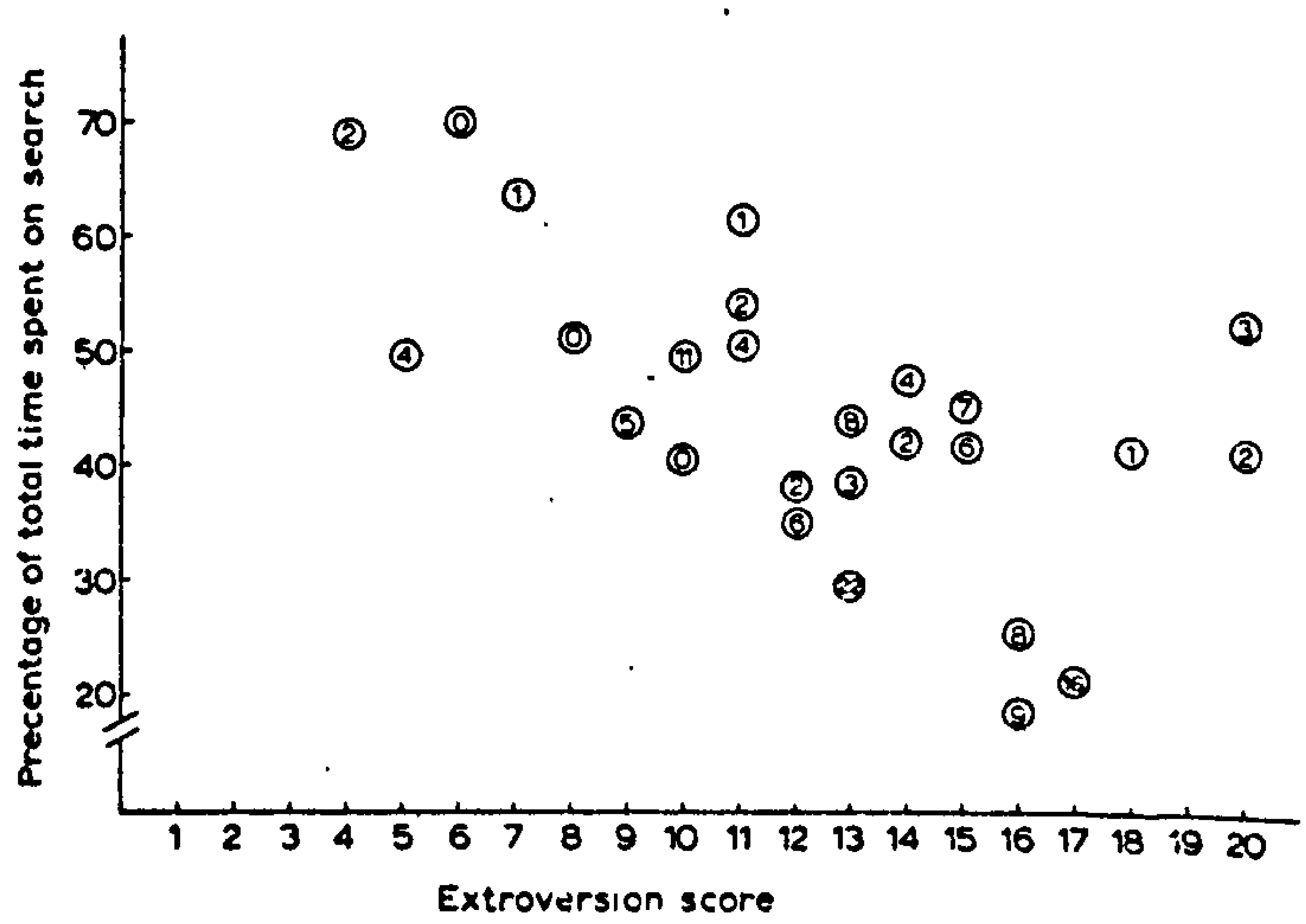


FIGURE 5a. Extroversion and search time on the Maze Test. (Digits show number of decision changes made by each subject.)

Taken from Jones and Wineman (1973)

however, on some ($\frac{3}{8}$) of the tasks N showed a significant curvelinear relationship with high and low N scores being faster than the mid N group.

In general, Russell asserted that in his data he was unable to find support for the hypothesis that extraverts work quickly, sacrificing accuracy, and that introverts tend to be slow and accurate. In fact, he noted a tendency for effects to operate in the opposite direction for the majority of Motor Tasks.

Although Russell's study produced findings unanticipated by Eysenck's theory, it was probably an inappropriate test of the various predictions. The male subjects were higher on N than would be expected on the basis of the standardization data and whereas E and N correlated -.15 for males, the correlation for females was -0.314, a relationship not found in the normative sample. Plots of E and N also revealed curvelinear relationships in the data. Russell was aware of these problems and attempted to cope with them statistically. Also Russell did not take into account the fact that the Nufferno Letter Series Test might have been too difficult for his subjects. This test is not normally administered to children as young as this, and given that they were of average intelligence, the Letter Series must have been particularly difficult. As Furneaux (1955) points out in the test manual - "The criterion situation can be taken to be Grammar School examinations of about G.C.E (Ordinary) standard. The test has,however, been used successfully with intelligent children down to about 11 years" (p L/11). If anything, Russell's subjects were slightly below average in intelligence (female mean 98,s.d 14.8;male mean 95,s.d 13.4).

Jones and Wineman (1973) used an on-line computer presentation of the Perceptual Mazes Test in a study of the relationship between extraversion,speed and accuracy. Difficulty thresholds were determined for each subject and they were then given speed-stressed and accuracy stressed instructions. The on-line task presentation enabled them to record the different amounts of time spent by subjects on different parts of the pattern. Thus their records enabled them to measure the amount of time spent on planning a route through the maze as well as other decision times and decision changes. Their study, based on the data for a group of 26 "relatively homogeneous" young adults showed a relationship between extroversion and speed preference. As extroversion score increased, there was a greater tendency towards faster responding and an increase in careless errors. No formal statistical analyses were reported in this study, but the trend is readily apparent in the accompanying diagram (Fig.5a).

The studies presented in this section have examined a variety of relationships between speed and personality on measures of intelligence or on narrow ability tests. Most of the studies cited have used the Eysenckian and on occasion, the Cattellian dimensions. There is also a large number of studies which have looked at the relationship between 'anxiety' and performance on tests.

Various questionnaires have been devised to measure 'anxiety'. The Taylor Scale (TMAS) was discussed in an earlier section. Mandler and Sarason (1952) have developed the Test Anxiety Questionnaire as a measure of the anxiety experienced by adults in test-taking situations. Similar questionnaires have been devised for adolescents and children (Mandler and Cowan 1958; Sarason et al. 1960). The test devised by Alpert and Haber (1960) unlike it's predecessors, attempts to identify testees whose anxiety in test situations either facilitates or disrupts performance.

Speilberger et al. (1971) distinguish between 'state' and 'trait' anxiety, the former being a transitory phenomenon, the latter equivalent to an enduring personality trait. They have developed a scale that aims to measure both aspects in test situations (The State-Trait Anxiety Inventory). None of these measures has been fully explored in relation to E and N although, as noted earlier, the TMAS appears to be a measure of both (Eysenck 1973b). Fleishman and Ellison (1969) for example, have found a correlation of 0.77 between the TMAS and the M.P.I. N scale.

Gaudry and Spielberger (1971) in their review of anxiety and educational achievement note that despite the different measures used for anxiety and intelligence, a low negative correlation is consistently found.

In a recent review, Entwistle (1972) has examined the relationship between personality (measured on the Eysenck and Cattell Scales) and attainment. His assessment suggested that an age effect may be important in that academic success at primary school is linked to stable extroversion whereas success at university tends to be associated with introversion. He does however caution that this apparently simple pattern may be complicated by a variety of interactions depending on intelligence level, type of institution and subject being studied.

A similar inversion emerges in with N, although for this dimension the strength and pattern fof relationship is somewhat different. At younger ages, N tends to show low negative correlations with achievement but at older ages the correlation tends to be low and positive. Exceptions have been found, and again other factors may be responsible for the observed pattern.

Naylor (1972) has also recently reviewed the research literature on the relationship between personality and educational achievement. He concludes that "general assertions concerning the relation between personality and academic achievement are unlikely to have general validity" (p.71). Sex differences in patterns of relationship as well as variation associated with school settings, teaching procedures and other factors are among the variables that lead to the uncertainty.

Anthony (1973) has attempted to account for the inversion of the relationship between E and IQ from pre-to post-puberty by proposing a differential developmental course for each variable. He states that E increases to a peak at 13 to 14 years and then begins to decline. General ability however increases until the twenties. On the assumption that individuals who are advanced (or behind) on one aspect are similarly advanced (or behind) on the other, the inversion can be explained. Unfortunately, while this is an attractive explanation, its validity is cast into doubt by it being an attempt to base notions of development on predominantly cross-sectional data (Schaie 1970). While there is some evidence that the E,N and L have decreasing reliabilities as retest interval increases (S.B.G.Eysenck and H.J.Eysenck 1973), there has been no extended attempt to follow the developmental sequence of these dimensions using an appropriate design. As Schaie (1970) emphasises, it "..has been known for some time that data on age-related phenomena will differ markedly depending upon whether it has been collected by means of cross-sectional or longitudinal research strategies".

In a recent study of over 2,000 children aged between 9 and 13 years (Whites,Negroes and Mexicans),Jensen (1973) found that the E,N and L Scales of the Junior Eysenck Personality Inventory showed low but significant and systematic correlations with achievement. Extraversion also correlated positively with three measures of IQ (Lorge-Thorndike Verbal and Non-verbal IQ and Matrices IQ). However, this was found predominantly in the 6th grade group, the highest value being .37 and the pattern across ethnic groups was not consistent. Even fewer of the correlations with N and L were significant and the same inconsistent pattern was shown. E,N and L made negligible independent contributions to achievement, but when combined in a multiple regression equation they accounted for a small share of the achievement variance. No attempt was made to examine relationships separately by sex.

The studies reviewed in this section are diverse in terms of the subjects tested, the measures used to quantify personality and intelligence, the sex and age of the subjects, and in terms of the quality of the research. There seems to be little doubt that personality factors correlate with test performance and that E and N have differential relationships with test scores, although the correlations are low and not always consistent, either from study to study or with the predictions based on Eysenck's theory.

9. Continuance, Accuracy and Test Performance

Unlike speed, the two other major components of Furneaux's model, continuance and accuracy, are not examined in detail in Furneaux's publications. Their role in his model and their interaction with speed will be discussed in Section B. The present chapter will focus on theory and research concerned with continuance and accuracy.

Furneaux (1961) proposes an individual difference dimension which is the tendency not to abandon attempts to find a solution to test problems. In empirical work, this disposition is indexed by the time elapsing between the presentation of an item and the recording of a "give up" response. This time interval is termed abandonment time and for any given individual there is presumed to be a distribution of such times for a set of items.

The disposition not to abandon items is termed "continuance", a term which he considers to be somewhat more general than the common term for this disposition, persistence. Continuance, as indexed by abandonment times, is considered by Furneaux to be partly determined by persistence, but .....

> "the decision to abandon an item may sometimes be made on grounds which involve an intelligent assessment of the effects on score of attempting a lot of items rather than persisting with a few". (Furneaux 1961).

The concept "continuance" has not been generally adopted in the psychological literature. In fact, the term persistence has been retained and appears to be used in a somewhat confusing manner. In its original form, persistence is a dispositional term, and refers to a personality trait. In the general literature (e.g. Feather 1962), "persistence" is also used to refer to the amount of time an individual actually worked on a difficult or insoluble task, or the number of attempts to solve the problem. For reasons which will become apparent, it is worth drawing a distinction between dispositional persistence and behavioural persistence. The latter is indexed directly by the time or number of trials to abandonment, whereas the former is inferred from abandonments. "Continuance" as used by Furneaux differs somewhat from both these conceptions. As will be seen, it is possibly the most appropriate term to use. Unfortunately, Furneaux never elaborated his views on the nature and determinants of continuance.

"Persistence" has been the subject of a number of major reviews (Ryans 1939, Eysenck 1960 b, Feather 1962).

In his review, Eysenck (1960 b) focussed on the evidence for "dispositional persistence" or persistence as a personality trait. On the basis of his assessment he concluded that persistence is a trait that is of a "relatively unitary nature", but that in addition, there are sub-groups of activities which cluster together to form more specific types of persistence, physical and ideational persistence being the two most important of these. As a trait it tends according to the evidence presented by Eysenck, to show slight correlations with intelligence, but "more impressive ones with 'w' or lack of neuroticism, and with introversion". (p.80).

Eysenck (1957) accounts for the differences in persistence between introverts and extraverts on the basis of differences in inhibitory potential. The extravert is assumed to generate excitatory potential slowly and at a weak level. Inhibitory potential is generated strongly and rapidly, but then dissipates slowly. The converse applies to introverts. It is proposed that on tasks requiring sustained or strenuous effort, the tendency of the extravert to generate strong inhibitory potential would lead to a lesser persistence relative to introverts (see Eysenck 1947; 1967 b).

Subsequent to Eysenck's (1960 b) review, little research appears to have been conducted on trait persistence, apart from the work of Thomas et al. (1968) on temperamental characteristics in infants and young children. These authors, in the context of a longitudinal study of individuality in children, have identified persistence as one of nine temperamental characteristics. They define persistence in terms of continuation of behaviour despite obstacles and on the basis of their data, suggest that persistence (as well as the other traits), can be identified in early infancy and has a degree of stability over time.

Feather's (1962) review, while accepting the evidence for trait persistence, is primarily concerned to develop an approach in which trait persistence is regarded as one of several determinants of behavioural persistence. His review is confined to a somewhat narrow definition of persistent behaviour. It focusses on the paradigm in which the individual is confronted by a very difficult or insoluble task and where no restrictions are placed on the amount of time or the number of attempts allowed. This definition does not take into account "easy" tasks of long duration, and "time allowed" is defined solely in the experimenter's terms: it does not take account of subject's perceptions of the amount of time. For example, subjects may abandon tasks for such trivial reasons as test sessions which extend into dinner breaks.

Studies of persistence, according to Feather (1962) cluster into three fairly distinct classes:

i. Those in which persistence is considered primarily as an enduring and pervasive personality trait which leads to predictable behaviour. Situational factors are given only minimal weight.

ii. Studies concerned with resistance to extinction in which persistence is indirectly implicated. Such studies, according to Feather, rarely consider persistence as an individual characteristic.

iii. Studies in which persistence is treated as a motivational phenomenon. In this group, the weight given to personality and situational factors can vary, but both are taken into account.

A study carried out by Feather and summarised in his 1962 paper, provides evidence for the inadequacy of a predominantly trait approach to behavioural persistence. This study will be described shortly. Feather's assessment of the research which conceptualises persistence solely in terms of resistance to extinction also draws attention to the inadequacies of such a view. Such studies emphasise the situational factors in persistence with little attempt to take into account personality variables. As he notes, they do not account for the variations between individuals in the same task environment. Studies have found that if subjects develop a high level of expectation of reward during the acquisition trials they tend to show low persistence when placed on an extinction schedule. Conversely, low expectation of reinforcement will tend to be associated with high levels of persistence. However, the generality of these observations is restricted by other factors. Feather (1962) points out that in partial reinforcement studies, an important element is the subject's perception of the control he has in the task situation: persistence is lowered when the subject sees the presentation of reward as outside of his control but is raised when the subject regards the reinforcement as dependent on his skill.

The series of studies which adopt, implicitly or explicitly, an interactional approach have served to underline the complexities involved. The basic model for such studies is concerned with the total motivation of the subject to perform the task in competition with the total motivation to perform an alternative. This model is somewhat limited in that the usual paradigm is concerned with two alternatives whereas in practice there may well be a fluid heirarchy of alternatives. The total motivation in any one direction is construed in terms of a number of sub-components each of which has to be determined for prediction purposes. It is then assumed that these components combine additively or multiplicatively together with any disposition in the individual to produce the individual variations in persistent behaviour on the specific task. The model is further complicated by the introduction, for each of the motivational components, three additional factors, motive strength, level of expectation and the magnitude of the incentive value (Feather 1962). A final element introduced by Feather relates to certain characteristics in the task situation, namely the extent to which the subject sees the outcome as being dependent on his own skill.

Feather's (1962) review of studies of persistence, and his own investigations were undertaken in the framework of Atkinson's theory of achievement motivation. These studies were aimed at testing a somewhat complicated set of predictions about persistence under different levels of achievement motivation, motivation to avoid failure, expectations as to task difficulty, and actual task difficulty (soluble and insoluble). Projective measures were used to define need achievement, and the Mandler-Sarason Test Anxiety Questionnaire to provide a measure of anxiety. It was then assumed that subjects with high projective test scores and low anxiety scores, for example, were individuals in whom the motive to succeed was greater than the motive to avoid failure. Four tasks, two of which were insoluble, were given to sub-groups of psychology undergraduates who showed varying patterns of scores on the two measures. Expectancy of success/failure was manipulated by giving subjects fictitious norms for each of the tasks. Persistence was measured by recording the amount of time or the number of trials on the insoluble tasks. Although not all predictions were supported, it was possible to show fairly strongly for example, that subjects in whom it is assumed that the hypothesised motive to achieve success is stronger than the motive to avoid failure persist longer at the initial achievement task when it is presented to them as easy rather than as very difficult. By contrast, subjects in whom it is assumed that the motive to avoid failure is stronger than the motive to achieve success do just the reverse and, according to Feather's data, persist longer when the task is presented as very difficult rather than as easy.

The results of Feather's studies have a number of important implications. Only those of immediate concern are considered here.

Although Feather made no attempt to measure trait persistence in his subjects it seems improbable that such a measure would by itself be sufficient to generate the diverse predictions that could be deduced from the Atkinson model. If nothing else, this study strongly suggests that there is no one-to-one correspondence between the "trait" and the performance: it would appear that behavioural persistence is at least a function of expectancy of success, and possibly many other influences as well.

It also seems fairly obvious, to the present writer at least, that Eysenck's theory of personality does not provide an adequate basis for predictions of the type tested by Feather.

Brody(1972) has also used Feather's study to clearly demonstrate that Cattell's approach, which attempts to unite source traits and situational variables, also cannot cope with the outcomes of the Feather experiment (see p. 22 to 25).

Finally, the results of Feather's study suggest that Furneaux was right to use the term continuance rather than persistence, a conceptual distinction which White (1973b) unfortunately has not retained. However, Furneaux's treatment of the topic of continuance does not readily lend itself to further characterization.

Since Feather's review, a number of studies directly or indirectly concerned with persistence have been published. These studies appear to have introduced a proliferation of motives and inconsistent findings related to behavioural persistence.

Di Ciaula et al. (1968) found some evidence in support of an intermittent reinforcement interpretation of temporal persistence in undergraduate psychology students on a set of anagrams. However, they were unable to account for the "somewhat surprising" finding that subjects who were given fictitious norms for the problems persisted longer than subjects who had had practice sessions on similar tasks prior to the persistence measures being taken. They also found no relationship between achievement motivation and persistence.

Stuempfig and Maehr (1970) divided their 84 sixteen-to eighteen-year old subjects into those whose conceptual development could be characterised as concrete and those who had developed to an abstract level. "Concrete" subjects persisted more (in terms of number of completed tasks) when given personal as opposed to impersonal feedback. The type of feedback had no effect on the "abstract" subjects. Clarke (1972) found that in his group of 40 twelfth grade students randomly assigned to feedback or no feedback groups, those who received feedback persisted more than those who were not given feedback. Those of his subjects who were high on achievement motivation but low on "affiliation motivation" persisted longer than other subjects. The least persistent were those who were high on achievement motivation but low on affiliation motivation.

Chaikin (1971) studied the effects of increasing success, decreasing success, constant success and constant failure in 30 male and 30 female undergraduates. Each subject was randomly assigned to one of the four groups and required to work on problem solving tasks. The group that was constantly successful showed a sharp drop in their desire to continue over time: the group tested under increasing success conditions showed the greatest persistence. The remaining groups showed intermediate levels of persistence.

The study by Shepel and James (1973) examined the effects of several variables on both temporal persistence and number of trials before abandonment. Their study included measures of "locus of control" and assessments of the extent to which performance on the tasks depends on the exercise of skill. Some tasks are heavily dependent on skill, others entirely on chance and yet others intermediate. The "locus of control" variable was introduced following the suggestion of Throop and MacDonald (1971), that individuals vary in the extent to which they believe life events to be beyond their control. "Locus of control" is proposed as a personality dimension. Subjects in the Shepel and James (1973) study were also given tasks under 100% and $33\frac{1}{3}$% reinforcement. The criterion measures were based on their persistence scores on an insoluble task. For their 96 student subjects (males) they found for example that for both measures (time and trials) under skill conditions there was a reversal of the partial reinforcement effect (i.e. less resistance to extinction) whereas under chance conditions, the partial reinforcement effect occurred. It was also observed that both internal and external locus of control subjects showed their maximum persistence when the locus of control and the task situation were congruent.

However, possibly the most significant finding in this study was a second order interaction for the temporal but not for the trials to extinction data. This finding (and others) is used by the authors to suggest that the earlier studies of persistence need to be "interpreted more cautiously."

One recent study (Gupta, 1973) using the MPI to obtain measures of introversion and extraversion, in addition to showing the importance of "motivation" in task persistence also indicated the differential effects of high and low motivation on introverts but not extraverts. Gupta measured temporal persistence on an arithmetic task in 40 applicants to a college and compared their performance with that of 40 students already accepted. He found that introverted applicants showed increased persistence scores on repeated testing: the already accepted introverts showed no such changes.

Studies of behavioural persistence in children have served to further complicate the discussion of it's determinants.

Barton and Barnard (1972) studied 40 children (4th graders) with high verbal and low spatial abilities and 40 with low verbal and high spatial abilities (measured on Thurstone's Primary Mental Abilities Test) under conditions of social reinforcement and no social reinforcement. The persistence measures were time spent on two tasks. Under conditions of social reinforcement, the high verbal-low spatial subjects persisted longer on the tasks than did children with the opposite ability pattern. This latter group of high spatial-low verbal ability also spent less time on the persistence tasks when socially reinforced than they did when no reinforcement was given.

Means et al.(1973) have found that middle class children persisted longer than lower socio-economic groups on a "very difficult jigsaw puzzle". They also found that subjects receiving no social reinforcement persisted longer than children who were either positively or negatively reinforced. The 90 children in this study had been randomly allocated to reinforcement conditions.

"Need achievement" has also been studied in children. Ollendick (1974) found that children high in "need achievement" persisted longer than did those who obtained low scores on the achievement measure. A correlation of 0.49 was found between need achievement and time on the insoluble task.

"Mastery motivation" was recently added to the list of motives implicated in persistence (Harter 1975). This form of motivation is regarded as the motivation which derives from the desire to solve a problem simply for the sake of discovering the solution. Harter (1975) included an insoluble problem in her study of 4 and 10-year old children, and also obtained measures of need for approval. She found that mastery motivation was particularly important, especially for the older boys, whereas need for approval was important for the girls but not for the boys. The younger children appeared not to be influenced by approval and they exhibited a special form of mastery motivation which was primarily concerned with the repetition of interesting stimulus events.

The concept of "accuracy" within Furneaux's model has a specific connotation, best illustrated by his hypothesised "problem-solving mechanism". Furneaux postulates several components to this mechanism including

a set of possible solutions to a given problem, the solutions being generated internally. With a possible solution available, it is transferred to a "comparator", a device which serves several functions. The possible solution is examined by the comparator and if found to be acceptable, a signal is issued which eventually leads to the individual making the overt behavioural response. If the comparator rejects the solution, it then re-instigates a search for a different solution and the cycle is repeated until continuance is exhausted or a satisfactory solution is found. The comparator is given certain characteristics in the mechanism, one of which is analogous to "bandwidth" in electrical and mechanical filters. Individual variations in bandwidth account for individual differences in error-proneness in the sense that broad bandwidth comparators are more prone to accept "nearly correct" solutions than are those of very narrow bandwidth.

The errors that Furneaux is concerned with are not so much those that arise due to lack of knowledge but rather those that arise from the "carelessness" or low standards of the system. Furneaux's mechanism (which incidentally includes a "persistence" device) is, apart from the language, very similar in conception to contemporaneous as well as more recent mechanisms. For example the TOTE (Test-Operate-Test-Exit) model of Miller, Galanter and Pribram (1960) has a "Test" component which not only detects the unsolved nature of the problem, but also checks if the solution returned by the "operator" is acceptable. This problem solver also has a "stop order" system, equivalent in many respects to Furneaux's continuance mechanism which prevents the system from going into an infinite loop. The Newell and Simon (1972) "Information Processing System", described earlier (IV i) also has many similar features (see Fig. 3). The nature of the accuracy checking mechanisms postulated by these various writers can ultimately only be inferred from studies of behaviour. The empirical task is one of delineating the sources of error.

Although there have been a number of speculations concerning the sources of error in cognitive tests, Payne (1973) points out that little is known about the causes of error or the nature of the checking process in solution verification. A number of researchers have resorted to the use of introspective accounts as well as the use of direct reporting by subjects of their ongoing thoughts during problem solving (e.g. Newell and Simon (1972). Errors can obviously arise in a variety of ways, the number depending on how finely one discriminates the component processes involved in problem solving. Payne (1973) suggests that the sources of error can include an inaccurate perception of the data of the task, a misinterpretation of the task instructions, and the complexity of the problem. Other factors, such as the method of presentation, could also be implicated. These sources are likely to be influenced by the implicit or explicit "costs and payoffs" inherent in the testing situation (Edwards 1961) as well as by what are regarded as the more orectic determinants, such as impulsiveness and the like.

Much of the research on error appears within two as yet unconnected areas of research. The first is in the area of comparatively simple classes of responses, such as accuracy on vigilance tasks and in reaction time, the second in the area of "cognitive style".

Broadbent (1971) has assembled much of the evidence in support of, or contrary to, the various theories devised to encompass performance on vigilance and reaction time tasks. Although a variety of theories have been proposed to account for the phenomenon of vigilance, and while these have focussed on performance decrement rather than on success, it is recognised that no single theory has been satisfactory. The theories have included arousal, expectancy as well as inhibition as major mechanisms framed within information, signal detection and decision theory contexts. Theories concerned with reaction times have attempted to encompass error as well as correct performance. They have been at least as unsuccessful as theories of vigilance in accounting for performance (Broadbent 1971) although one important finding of relevance to the present discussion has emerged. Broadbent (1971) notes that the speed and accuracy of decision, as seen in reaction time tasks, are influenced by what he calls "the general characteristics of the process" (p.323). The quality of the performance depends to some extent on the criteria which are imposed on the decision. If, for example, the subject operates with criteria that require lower levels of evidence, faster levels of performance will arise at the cost of accuracy. As Broadbent (1971) states:

> "Both speed and accuracy will be sensitive to general changes of criterion placement".

Factor analytic studies of cognitive tests have yielded factors identified as accuracy components.(e.g. Howie (1956) and Mangan (1959) described earlier). Both Howie and Mangan reported bipolar factors which contrasted speed with accuracy on tasks of low conventional difficulty. A perceptual speed and accuracy factor was also found in the more recent study reported by Lohnes (1966). Vernon (1961), in his book on the structure of human abilities concluded that speed-accuracy shows a "..... fair degree of generality ... when the material is easy and time restricted." (p.152).

One trait, impulsiveness, appears to be directly relevant to accuracy. Impulsiveness is an important component of extraversion (Eysenck 1975) and there is some evidence that extraverts make more errors on the Nufferno speed and level tests, for example, than do introverts (Payne 1973).

There has been a substantial amount of research in recent years on the behavioural dimension in children termed "reflection-impulsivity". Although it is not certain that the impulsivity component of extraversion is in any way related to the impulsivity dimension proposed by Kagan and his colleagues (e.g. Kagan and Kogan 1970), there are sufficient similarities to warrant a brief examination of the reflection-impulsivity literature in the context of a discussion of accuracy and its possible personality accompaniments. For example, Kagan and Kogan (1970) state that:

> "The reflection-impulsivity dimension is concerned with the degree to which the subject <u>reflects on the validity</u> of his solution hypothesis in problems that contain response uncertainty."

In a recent paper, Kagan and Messer (1975) note that both latency and number of errors on the Matching Familiar Figures Test (MFF) are used to classify children as reflective or impulsive. This procedure was introduced so that they could differentiate between children whose fast response times were associated with many errors and those whose fast response times were associated with errorless performance.

The paper by Kagan and Messer (1975) was issued in response to a critical analysis of the MFF as a measure of reflection-impulsivity by Block et al.(1974) who concluded that the MFF is an inadequate procedure for measuring this dimension. The attempt by Kagan and Messer (1975) to counter these criticisms was rejected by Block et al.(1975), who persist in their view that time is unimportant as far as personality implications are concerned. Instead, they assert that the major differences arise in association with accuracy-inaccuracy, that "Accurates" are :

> "ego-resilient individuals well stocked with assimilative structures but able also to accommodate to the previously unexperienced: the Inaccurates we see as comparatively brittle individuals, more rigid, less resourceful and therefore more susceptible to anxiety."

A major limitation of many of the studies examining differences between reflective and impulsive children has been their failure to control for IQ, either statistically or by matching (Block et al.1975). However, a study by Achenbach and Weisz (1975) suggests that IQ may not be the most important variable. They found that when Mental Age, as an index of developmental level, was partialled out of their data, there was virtually no evidence for the existence of an independent trait of reflection - impulsivity as a component of "cognitive style". If their conclusions are also found to apply to older children, then, obviously, the vast amount of research on this trait will have to be re-evaluated. As these authors used the KRISP (Wright 1971) and only one item of the MFF, further study is needed to assess the generality of their findings across tests as well.

On the basis of the preceeding paragraphs, it can be seen that Furneaux's choice of the term continuance, rather than persistence, was justifiable although it is doubted that he could have foreseen the complexity with which it is now endowed. There appears to be little doubt that behavioural persistence is complexly determined: what it's precise determinants are remains to be specified. In a sense, much the same could be said about accuracy. Although Furneaux appeared to refer to some central mechanism for error checking, such a mechanism is at present inaccessible. At the behavioural level accuracy could be a function of several factors, at least some of which may be personality linked.

10. Concluding Remarks: Speed, Continuance and Accuracy

The concepts "speed", "accuracy" and "continuance" appear to be uncomplicated. On a closer examination of their theoretical characterisation and their transformations into operational variables the simplicity disappears. The complexities which emerge are confounded further by the attempt to integrate them within a personality theory. It is unfortunate that no attempts have been made by psychologists to clarify the basic concepts involved: even Furneaux's "logical atomisation" of his primary concepts appears to be insufficient, One of the problems appears to be what Harré and Secord (1972, p.36) call "the conceptual leap from the theory to the operation ..... over-emphasizing empiricism at the expense of conceptualization ..." To the present writer at least, this would appear to be one of the major problems which emerged from a closer inspection of the research literature. This is particularly the case with the study of speed and one of the many ramifications of this is the ensuing empirical confusion.

While it may still be difficult to isolate strong empirical conclusions from the foregoing research, there is little doubt that the methodological and technical gains have been substantial. Unless they are of immediate interest, age effects must be partialled out of the data. Even when IQs are used, their correlation with age needs to be examined in the specific sample. The fact that IQ's are "age corrected" does not mean that they won't show correlations with age in a given sample. There is now sufficient evidence of sex differences in a variety of psychological characteristics and sex effects need to be properly controlled. It has also been suggested that neither group testing not crude timing procedures yield data that are likely to be of sufficient accuracy and careful attention has to be given to these procedures in research. Instructions have to be clearly stated and caution exercised that instructions, setting and aims do not produce conflicting reactions either within subjects (such as stating that there is no need to rush yet using overt timing procedures) or between subjects, such as might arise when subjects are free to choose which part of the instructions they will relate to their own performance.

No study is likely yet to satisfy all the requirements of a methodologically ideal investigation and the present study is no exception. The author has attempted to take account of what guidance there is in the literature but deficiencies remain. It is hoped that these will be explicit in the presentation which follows.

The preceding reviews of the literature identify many theoretical and technical problems. Those selected for the present investigation are specified later as the hypotheses which have determined the research to be described in subsequent chapters. They are based directly on the theory proposed by Eysenck and with some exceptions, attempt to replicate research findings reported in a number of studies.

In examining personality performance relationships, an important decision that needs to be taken at the outset is whether or not speed is to be emphasised in both instructions and,indirectly, in the way in which the experiment is conducted. Almost all of the research considered in the review has been concerned with stressed speed. That is, even when instructions have given subjects specific options, the use of group testing and overt timing must have had the effect of stressing speed. Hence a major research question is how personality and performance are related when neither the instructions nor the other procedures give the impression that speedy responses are required. The present study was designed in part to examine this question.

SECTION B.

The preceeding chapters have been concerned with introducing this study, and providing it with a general framework. This framework is composed of several elements, encompassing the limitations of conventional tests, some of their central problems and some of the more recent approaches to the fundamental problems of intelligence measurement, particularly the problem of difficulty. Subsequent chapters examined certain theoretical approaches and the empirical studies specifically concerned with speed and its relationship with intelligence and personality.

Section B of this thesis is concerned with more specific aspects, particularly the work of Furneaux and its incorporation into the broader context of Eysenck's model of personality and the theory which relates personality to test performance.

VI. FURNEAUX'S ANALYSIS OF A FORMALIZED PROBLEM SOLVER.

1. The Conceptual Model:

As indicated in the preceeding chapters, the work of Furneaux is central to Eysenck's (1967a, 1973a) approach to the measurement of intelligence. However, his work is also important because it appears to overcome the psychometric problem of item difficulty scaling. This latter aspect is the immediate concern of the next chapters.

Butcher's (1968) remark about the 'decided obscurity' of Furneaux's analysis applies to two sections of his work, the analysis of problem solving and the design of difficulty scales. The feature of his presentation which hampers ready translation is the mixture of abbreviations and logical and mathematical notation. To try and overcome some of the problems which this form of presentation entails, a procedure has been adopted of presenting all notation on the left-hand side of the page, except in those instances where it is essential to use it in the text. Reference to the notation is indicated by " * ".

Furneaux's approach is guided by two main principles. Firstly, the human problem-solver must be conceptualized as a "black-box", the input - output characteristics of which are to be specified by unambiguously defined observations. Secondly, in the initial stages, the aim should be to devdop an oversimplified model which has only those characteristics specifically assigned to it.

Thus, the human problem solver, is, in the first instance
PB — regarded simply as a problem-solving box * whose mechanism is
unspecified. The nature of this box is such however, that when
I, oe — a problem is supplied as input * an "essay" * or solution is
produced on the output side which is the attempt at solution. The
te — problem solving process requires time which can be measured * and
the solution offered can be evaluated as to whether or not it is
O_r, Ow — correct * or incorrect *. The elapsed time can then be defined as
tr, tw — time to correct solution * or incorrect solution * depending on the
nature of the solution.

The model next goes on to assume the repeated input of the same
pI, aPB — problem * to a particular problem box *. The output sequence is
qr — conceptualised as being unpredictable with some proportion correct *
qw — and the remainder incorrect *. It is further assumed that even
for the correct responses alone, the solution times vary between
Rt and (R_t+H_{tr}) — specifiable limits *. These solution times in effect define a
$\bar{t}_r$; V_{tr} — distribution with its own mean *, variance * and range. A similar

tw; Vtw; (Rw + Htw).

distribution can be defined for the incorrect solutions *.

For each individual problem box subjected to the repeated input of the same problem, the resulting distribution parameters are, as a set, symbolized by 'M'. There will be one M set specific to the correct solutions * and another to the incorrect solutions *.

p a Mtr

p a Mtw

Furneaux contends that the investigation of the problem-solving characteristics of a particular problem-box must begin with the collection of the "M-statistics" based on the repeated input of the particular problem. Similar data can then be obtained for other problem-boxes * based on the input of the same problem. With these statistics determined, it then becomes possible to make comparisons among the various problem boxes. Which particular statistic is used as the basis for comparison, is, in the absence of the requisite information, irrelevant. That is, no single statistic is inherently preferable to any other. Nor should one be used to substitute for any other in the absence of evidence that they are all highly intercorrelated. It would also be meaningless to, say, compare correct solution times from one problem box with the incorrect solution times of another. Furneaux also points out that it would be unwise to derive "some score which was a more or less undefined function of a lumping together of several of the statistics and then to imagine that it could have more than an accidental significance".

b PB, c PB --- etc.

The selection of problems to be included in any test, is, as Furneaux points out, a function of the purpose (theoretical or applied) of the test. Further, there is no inherent reason why one set derived from traditional scale construction procedures should be better than a different set chosen on the basis of M statistics. However, the task of selection requires that there be some acceptable means of classifying problems. It is at this point that the notions of 'difficulty' and 'type' have to be introduced.

Some of the limitations of the traditional concept of difficulty have already been discussed and after considering these, Furneaux reaches the conclusion that we have to accept that "problems can only be classified in terms of the differences in response characteristics which they evoke". Given that there is no reason as yet for selecting among the various M statistics, scaling procedures could be carried out using any of the M statistics considered relevant. There is no one statistic that is the best measure of difficulty and there is also no single meaning for difficulty. Furneaux's view of difficulty is therefore very similar to modern conceptions of reliability and validity, namely, the recognition that these are all generic concepts (Stanley 1971; Cronbach 1971).

The purpose of scaling, as Furneaux points out, is to enable some form of classification, and this consideration must always be taken into account in any scaling exercise. In view of their being no a priori basis for selecting any of the M-statistics, at this stage of the analysis it is necessary to define a separate scale for each. Collectively, such scales are called "D-scales". Specifically, a scale which is confined to mean correct solution times for the problem is signified as a D_{t_r} scale; the scale associated with proportion correct would be identified as a D_{q_r} scale, and so on.

The task of item scaling is illustrated by Furneaux with reference to scales associated with correct solution times. If we are given the scale positions of two items, p and q, and we know beforehand the mean correct solution time to problem p for a particular problem-box, then we should be able to predict with fair accuracy the mean correct solution time for problem q.

Symbolically, the requirement is stated as

$$_q t_r = f\left(q\, D_{t_r}, a K_{t_r}\right) \qquad ----- \; 1.$$

$$= f\,(D, K)\, t_r \qquad ----- \; 2.$$

where: $q\,D_{t_r}$ is the scale position of item q

$a\,K_{t_r}$ is an individual constant assigned to individual 'a' and is based on his mean time to correct solution for problem p.

Equation 2 is the general form of 1.

From the point of view of the model and its application, it is necessary that the form of equation 2 is constant across individuals whereas the value of K_{t_r} can vary from one person to the next. Once a suitable form for equation 1 and the scale positions of p and q have been found, then, according to Furneaux, it becomes possible to compare various 'problem boxes' in terms of their mean correct solution times to some problem. What is required from each is the value of K instead of all values of mean solution time for all possible problems.

While it would be desirable to have a common form to f(D,K)tr for all possible types of problem so that each could be allocated a position on a single D-scale, data from correlational studies of conventional tests suggests that this is unlikely. The data which indicate that this may not be possible are derived from the many factor-analytic studies (Vernon 1961) which reveal group factors in human abilities. Furneaux anticipates that sets of problems with a common form of f (D, K) t_r may be found, which he calls "types" which in turn may or may not correspond to group factors. At this stage of model development, the type-structure is not known. It is also possible that not all D-scales will be generalisable across all problem-boxes. It is necessary to contemplate the possibility that D_{t_r} scales would vary say, between schizophrenic and non-schizophrenic groups, or that different personalities might produce differences in the scales.

The major advantage of the procedure suggested by the model is, according to Furneaux, that problems of standardisation are avoided. They are replaced by the problem of determining the scale position of the items, but this task can be accomplished by the procedures he details for D-scale construction which will be described in the next chapter.

Up to this point, the model is unencumbered by the human characteristic generally known as 'persistence'. Human problem solvers give up their attepmts at solution if an answer is not forthcoming and go on to the next problem. Furneaux is careful to emphasise that persistence is not necessarily the sole determinant of abandonment. Newell and Simon (1972) have for example suggested that subjects may decide to abandon an item because their list of potential solutions has been tried and exhausted. The decision to abandon an item may also be based on the test-takers conclusion that rather than waste too much time on one item, he will have a better chance of a higher score if he tackles all the items.
C Furneaux therefore choses "continuance" * instead of persistence as it has no "aetiological presuppositions".

ST Continuance is added to the model by installing a time-switch *
PS and the new device given the name of a problem solver *.

The normal state of the time switch is "off" but it is activated
t_s immediately a problem is input. After a period of time * when continuance is exhausted the switch reverts to its normal state. The coupling between problem box and time switch is such that the switch only becomes activated on the input of a problem and the problem solving process is only initiated with the switch 'on'. The problem solving process is automatically terminated when the time switch reverts to "off" when continuance is exhausted. A further feature of the time switch is that it has no means of

distinguishing between different types of input. When continuance is exhausted before a solution is available, the output from the problem solver takes the form of a " -- " * on the answer sheet. However, when an answer is output, it has the effect of switching off the timing mechanism.

O_s

As with solution times, the abandonment times have a distribution with a mean *, variance * and range *. The sequence of abandonment times cannot be predicted. The set of statistics associated with the distribution of abandonment times generated by the uncoupled time switch is symbolised by M_{ts}.

$\bar{t}_s$, V_{ts}, S_t and ($S_t + H_{ts}$)

The coupled mechanism, on the input of a problem, initiates a "search" for a solution. This search is terminated by the production of an essay or when continuance is exhausted.

Special consideration needs to be given to the distributions of solution and abandonment times. It will be recalled that for the uncoupled problem box, on repeated input of a particular problem, a solution is always offered. This solution, the essay, can be either correct or incorrect. Associated with each essay is a response time. The distribution of response times can be divided into the distributions for correct and incorrect times. A distribution is also associated with the abandonment times when the time switch is uncoupled. Once the coupling takes place, the distributions of essay times and abandonment times become altered. If an essay arises from the coupled device, before continuance is exhausted, the abandonment time distribution in effect loses one time value which it would not have done had the time switch not been coupled. Hence the distribution of abandonment times becomes modied *.

$\dot{M}_{ts}$

In a similar way, if an abandonment occurs before an essay arises, the distribution of essay times loses a value that ~~it~~ would not have been lost had the problem box not been coupled to a time switch. It is therefore also necessary to distinguish a modified distribution * of essay times. The " . " notation is used by Furneaux to indicate the modified M statistics. Individual statistics that are changed are also indicated by the " . ". *

$\dot{M}_{te}$

$\dot{V}_{te}$, $\dot{V}_{ts}$, .. etc.

The introduction of continuance into the model considerably complicates the attempt to develop difficulty scales and the practical determination of the M statistics. To simplify the analysis, Furneaux assumed that all types of D scales are identical. This assumption was later subjected to experimental examination.

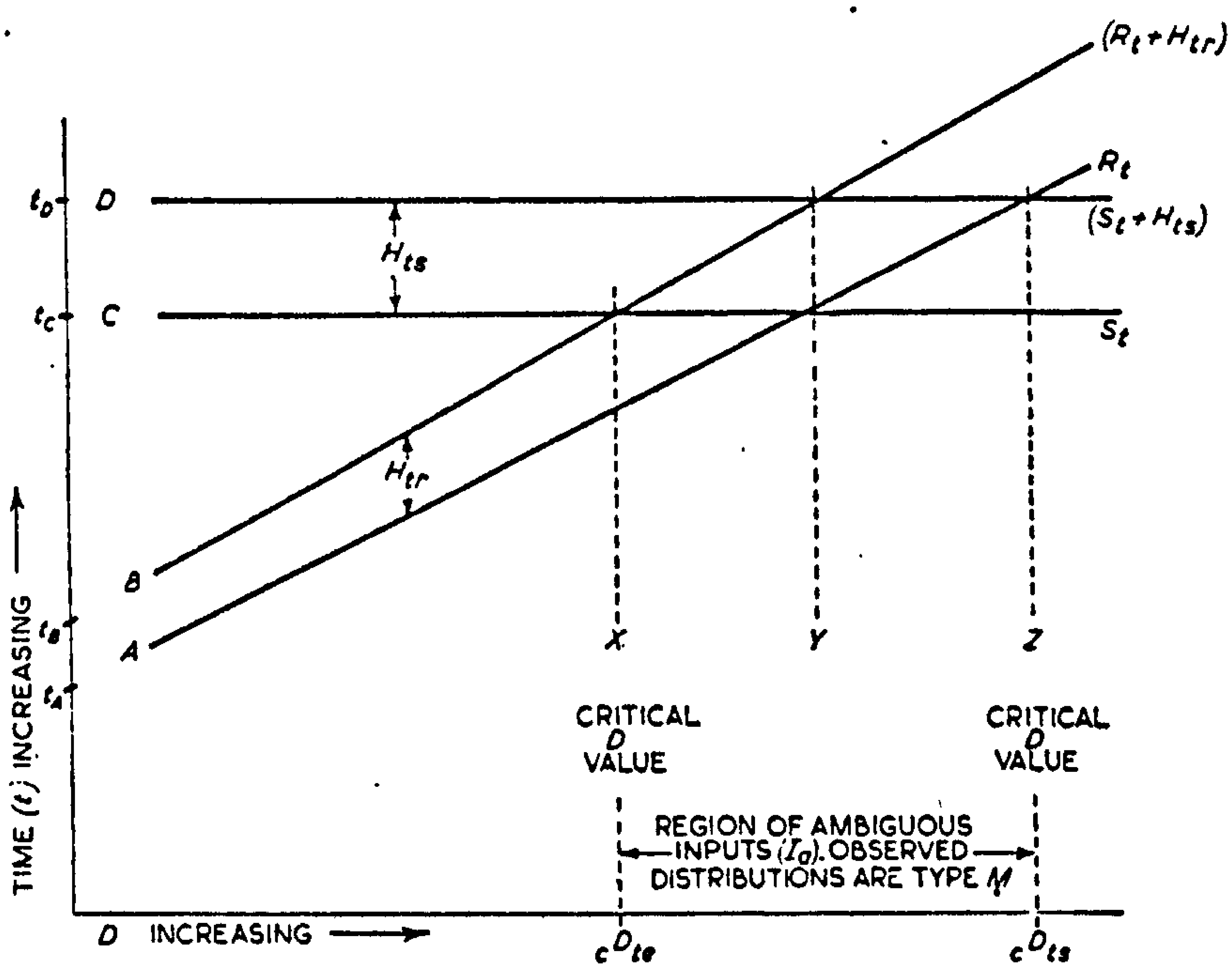


FIGURE 6. . The Interaction of Speed and Continuance, Theoretical
Taken from Furneaux (1961)

Fig. 6. has been taken from Furneaux (1961)
and will be used to illustrate the complications which arise when continuance is introduced.

Increments on the abscissa represent increments in difficulty (D). Increments on the ordinate represent increments in time. Line A is the lower bound of the range of solution times at increasing D values, line B, the upper bound for the uncoupled problem box. Lines C and D are the respective lower and upper bounds for the range of continuance times produced by the uncoupled time switch. Although the Figure assumes they are unrelated to D, the basic conclusions would not be altered if their slope varied (provided of course they did not coincide or go below A and B).
The pairs of lines intersect at the points indicated by the vertical lines X, Y and Z.

From this diagram, it will be apparent that at all D values
c Dte below X, * an essay will always arise. Performance is not interfered with even though the time switch is coupled. At all
c Dts D values above Z, * no essays will appear because the problem solver will always be switched off by the time switch. All responses at this point and beyond will be true abandonments.

In the region X to Y, two types of response will appear. These will be either essays or abandonments. Which will arise depends on the interaction between D and the problem solver. The number of essays having values near to line B will be reduced. This will occur because on a proportion of occasions, the time switch will switch off the action of the problem box, before an essay can be generated. However, all times near to line A will still emerge.

In the region Y to Z, a somewhat different set of consequences result. Any essays whose times are greater than maximum continuance (defined by line D), and which might otherwise have arisen, will not appear because of the operation of the time switch. Those that do so, succeed in emerging because they are close to line A, and it is this group that will predominate.

Thus, the effect of the time switch at D values beyond X is to change the distribution of essay times. The closer to Z, the more the distribution changes.

Furneaux refers to problems with D values lying between X and Z
Ia as <u>ambiguous-inputs</u> * for the particular problem solver. Strictly, as he points out, it is not the inputs that are ambiguous, rather the statistics generated by their outputs are ambiguous. Those

inputs which do not generate ambiguous outputs are known as
Iu unambiguous inputs *.

The region of ambiguous inputs is bounded by critical D-values. The lower bound marks the transition from
c Dte unambiguous to ambiguous region for essays * and the upper bound,
c Dts the transition to "unambiguous abandonments". * Below the lower bound, the M-statistics for essays are unaffected; above the upper bound, the M-statistics for abandonments are unaffected. Within the bounds there is a reciprocal loss of data from either essays to abandonments or abandonments to essays.

The human problem solver, following Furneaux's arguments, cannot be partitioned into a problem box and a time switch. Because of this, there may well be a wide range of D-values within which it is not possible to compare individuals. The location of the critical D-values is said to vary between individuals, so that a given D-value may constitute an ambiguous input for one problem solver and an unambiguous input for another. In the case of the former, the M-statistics will be distorted whereas this will not be so in the latter case. This would apply particularly in the case of tests made up of heterogeneous items.

On the basis of these considerations, Furneaux reaches a conclusion of major practical import, that tests intended for use with heterogeneous groups will have to be comprised of "easy" items if any of the statistics associated with essays are to be computed. The only alternative which he considers to be viable is to raise the mean abandonment time to such a level that the interaction between abandonment and difficulty no longer operates within the range of difficulty required. In terms of Fig. 6, this would mean a vertical shift of the lines C and D upwards. In practice, this would be accomplished, to some extent at least, by inducing a high motivation not to abandon items, and by discouraging attempts through individual testing, to move on to other items on the test sheet.

While the preceding paragraphs have related to the general concept of essays, a similar analysis could be made for either of the components of essays, namely, correct and incorrect solutions. In these cases, further discriminations need to be made and some new assumptions introduced.

It is possible that the f (D, K) which characterises correct response times will not be the same function which is generated when considering incorrect response times. It is also possible that the respective M-statistics will have differing values when computed. Further, an unambiguous input which would be

Ia (t_r) associated with a correct response * in one case may be

Ia (t_w) associated with an incorrect response in another *. Finally, there may be different critical D-values associated with

c Dtr; c Dtw correct * and incorrect * solution times.

Following a procedure similar to that presented when considering essays, some potentially correct response times will be lost due to the operation of the time switch, and will be replaced by abandonment times. With a knowledge of the form of f $(D,K)t_r$ and the M-statistics associated with correct solutions and abandonments, it is possible to compute the proportion of potentially correct responses that can arise at any specified difficulty value. This proportion will be unity for all difficulty values less than or equal to the critical D-value and zero at all D-values greater than or equal to the critical D-value for abandonments. The curve showing the approximate form of this relationship is presented in Fig. 7, as curve B, and is called the <u>completion characteristic</u>. It's abcissa represents increments of D; the ordinate represents the

Psr proportion of correct responses,* that arise out of those potentially possible.

While it might be thought possible to compute this proportion out of the ratio of correct to abandoned responses at successive D-values, Furneaux points out that this would be inappropriate. This is so because any abandoned item might well have been a wrong solution had the time switch not intervened.

It should be noted that the completion curve is, in effect the curve which represents the responses observed through the operation of the problem solver. It does not represent his 'ideal' response characteristics because of the confounding effects of the time switch. That is, some correct responses will be lost because the time switch (continuance) intervenes before the response emerges.

It is however possible to generate the hypothetical curve of performance of the problem box (i,e. problem solver without time switch) in terms of the proportion of correct responses generated without the encumberance of the time switch, at each D-value.

qr The curve relating this proportion * to difficulty is presented as curve A in Fig. 7. This curve is known as the <u>accuracy curve</u>.

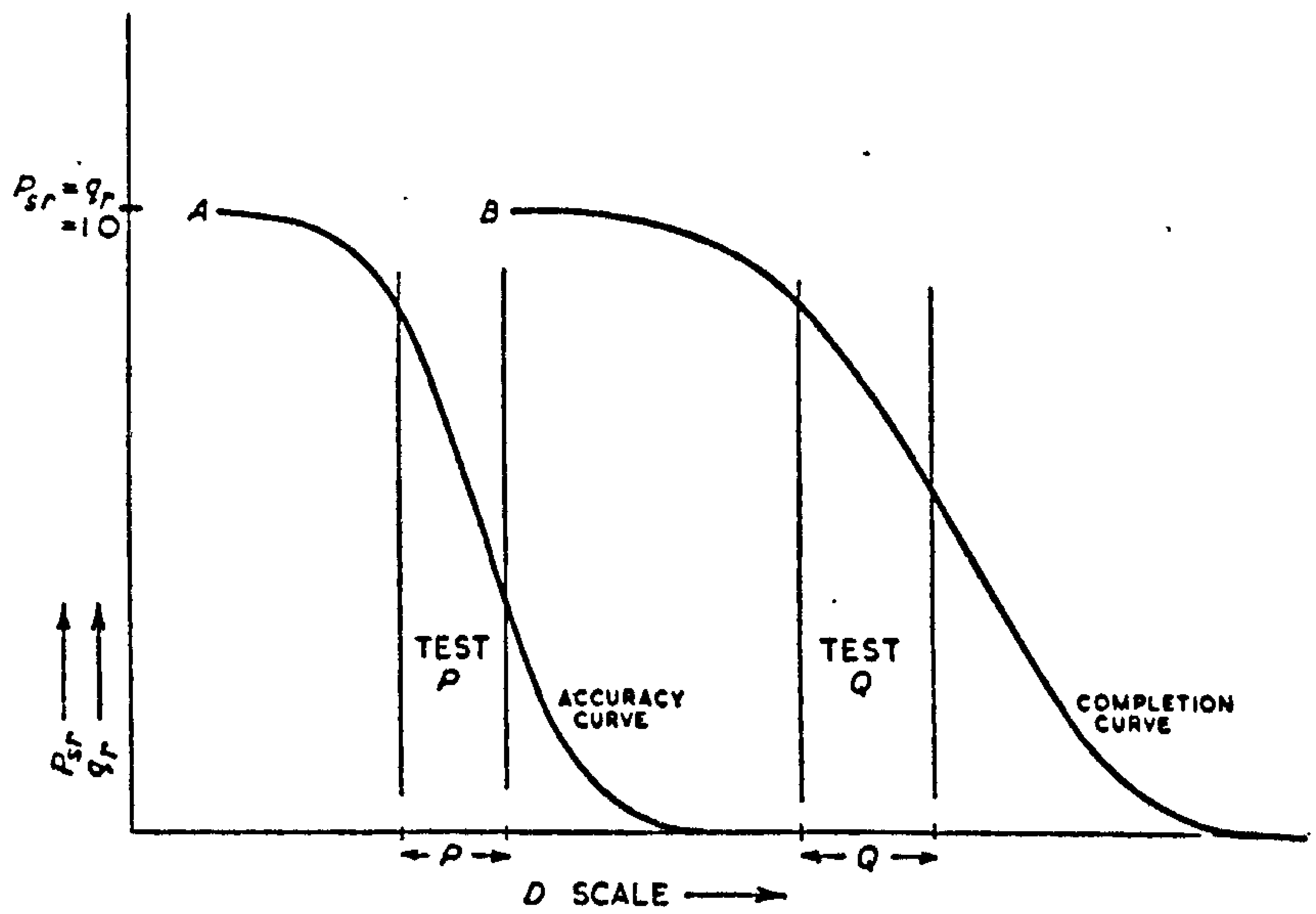


FIGURE 7. THE INTERACTION OF COMPLETION AND ACCURACY, THEORETICAL

Taken from Furneux (1961)

P_{sqr} K_{qr}

At any given value of D, the probability that a correct response will occur * is a product of two separate probabilities, P_{sr} and q_r. At this D value, this probability is a constant *. If the value of this constant is only moderately correlated with speed and continuance in a group of problem solvers, than, as Furneaux notes, for different problem solvers, the curves A and B can differ in both their "relative and actual positions to a very considerable degree".

The position of these curves for a particular individual and set of items determines to a great extent his performance on the test. For the set of items covering range P in Fig.7. the main determinant of success is accuracy. This individual is low in accuracy, as illustrated by the rapid fall off in proportion correct for small increments of difficulty, whereas, his completion - a function of "speed" and "continuance" - is much higher. However, the absence of overlap between curves A and B indicates that for test P, completion does not enter into his performance. On a test covering the range of Q, which is positioned in such a way that accuracy does not enter into performance, all the responses would be incorrect.

Underlying this analysis is Furneaux's fundamental point that at any given value of D, responses to a test can vary as a function of the accuracy and completion characteristics of the problem solver. What may appear to the outside observer as a standard input is received in a way which varies according to the characteristics of the problem solver. One cannot assume identical "internal processes" being elicited by the standard input. For example, the "same mental test, if designed and scored in the conventional manner" will measure varying combinations of speed, accuracy and continuance when given to different testees. The crucial point is the presence in the test of an item with a D-value which initiates the operation of continuance. In so far as these critical D-values differ for different problem solvers, the scores which they produce will be differentially determined.

Furneaux goes on to make a further fundamental point about the notion of the "speed free" test. One cannot assume that a test set without time limit is "speed free". As he says,

> "In any test which includes problems having D-values which for a particular subject, are greater than c D_{tr}, score is determined in part by the completion characteristic, which is itself, in part, a function of speed (K_{tr})-".

How does this analysis apply to the commonly encountered approach to testing, the approach which provides the foundation for theories of intelligence and other abilities?

In the simplest case, consider a test made up of "easy" items, that is items with low D-values which do not produce incorrect solutions for the group of testees. Such items will not involve continuance and under time limit conditions, the only factor influencing score will be speed. (At this point, we can disregard other psychological complications, such as test anxiety). If the testees have then to go on to a set of items of increased difficulty, a different set of factors comes into operation. The set of moderately difficult items is such that most members of the group produce some incorrect solutions but none of the items are so difficult as to lead testees to abandon them. In this compound test, set with a time limit, the slowest members will be working on the easy items whereas a faster group will have completed the easy items. In this latter group, members will have been working on the more difficult items and their scores will be partly a function of their speed, and partly a function of their tendency to be inaccurate.

If this test is further compounded by the introduction of very difficult items, then, for the fastest subjects, scores will vary as a function of individual continuance.

The already apparent complexity is further complicated by such factors as the degree of homogeneity in the group being tested. For a group with homogeneous abilities, a test could be tapping mainly speed, mainly accuracy or mainly continuance, depending on items difficulty. For heterogeneous groups, there will no longer be a preponderance. The factors determining score could be speed alone in some subjects and various combinations of determinants in others.

Although Furneaux's analysis has been restricted to a simple conception of the problem solver, it will be appreciated that it's implications are of major significance. These implications do not only extend into the use of test scores for developing our understanding of intelligence. Taken seriously, they are of relevance to the study of different groups (for example those with and without a recognised psychiatric disorder), the same group at different ages or in different settings, and so on. Furneaux himself suggests how his analysis may be relevant to the study of the development of ability. Thus a test at age 5 may, because it is "difficult" elicit one set of determinants. Reapplied 5 years later when there has been some development of intellectual abilities, the test may elicit a different set of score determinants. Until such time as it is demonstrated that the same determinants are operating in the same way and to the same extent at the two ages, it will be difficult to draw any meaningful conclusions about performance at these two ages.

From the point of view of initiating research into human intellectual function, Furneaux's analysis appears to set a more appropriate basis for such work. He has, in effect, defined some of the conditions necessary for the 'proper study' of speed, accuracy and continuance. To get any of the statistics associated with correct solution times, it is necessary to use data from items which have D values below the critical difficulty for the subjects. To examine the continuance characteristics, data should be used only if they are derived from items with D values above the critical level for continuance.

A further contribution of Furneaux's analysis is it's relevance for applied testing. At the very least, it suggests that test scores may not simply be a function of the abilities sampled by the test content. While applied psychologists do not in general have an oversimplified view of the determinants of test scores, the range of influences they consider is somewhat narrow even if relevant. A review of research on cognitive abnormalities by Payne (1961) provides one illustration of the relevance of Furneaux's analysis to the assessment of cognitive functions in psychiatric patients.

Despite these important implications, the impact of Furneaux's logical analysis is dependant to a large extent on the outcome of research which aims to test the viability of his views. In the next chapter, Furneaux's own empirical work is considered in some detail as it represents the first test of model.

2. Empirical Findings:

The present chapter will focus on the procedures and findings reported by Furneaux (1961). To avoid confusion, the detailed procedures for deriving D-scales are described in a later chapter.

As presented by Furneaux, the D-scale procedures are required to produce various constants and statistics that are only minimally correlated or uncorrelated. Further, it is necessary to demonstrate that whichever of the f (D,K) functions is derived, it should be sufficiently general for all, or as he puts it, "useful sets" of individuals.

In the major study, Furneaux used data from a group of 235 soldiers, ranging in age from 18 to 30 years (mean age 19.4 years). These subjects were selected from a larger group on the basis of their scores on the Dominoes Test. Only individuals scoring between -1.5 and 2.0 s.d's were selected. The lower cut-off was introduced so as to ensure that subjects would be able to follow complex test instructions. The upper bound was chosen because very few subjects scored beyond that point. Furneaux was thus able to obtain a fairly uniform distribution of scores.

The test materials consisted of items from the Letter Series of the Thurstone Primary Mental Abilities, arranged in 5 cycles. Each cycle contained items of varying difficulty, the difficulties being of the conventional type. These values had been computed from data available from a previous administration of the test. The items had to be worked through in a strict sequence and once abandoned or answered, could not be re-attempted.

The subjects were given instructions "designed to encourage high motivation and the evidence suggests that this was achieved". Neither the instructions nor the evidence are detailed. Testing was carried out in groups of about twenty subjects, with three hours of testing for each group spread over one day.

Solution times were measured to an accuracy of 2 seconds using a timing device which was described earlier. Each of the times was corrected so as to eliminate time spent in reading the device and recording times in the answer booklets. The correction factors were obtained on the same subjects in an earlier part of the study. No details of this are provided. Of the 235 items in the original set, 80 remained in the final set, the rest having been eliminated because they were not attempted by all subjects or because they were found to be unsuitable following a conventional item analysis. The reduced test and Dominoes scores correlated 0.84.

Furneaux reports having to eliminate segments of his data because of variations in the continuance times between early and later periods of testing. Failure to do this would have meant that a number of statistics, rather than a single index of continuance would have to be used, making D-scale construction much more complex. What is not clear is whether the discarded data led to the reduction of the total items used to 80 or whether the already reduced set of 80 was further reduced. Thus, it is not known how many items were used for scaling.

One of the major goals of Furneaux's work was to find a uniform function relating solution time to item difficulty and speed, i.e.

$$t = f\,(D,K).$$

The particular scaling exercise described in the 1961 paper is that relating correct solution times to speed and difficulty. The most convenient function for this purpose takes the linear form

$$y = m\,x + c$$

In Furneaux's notation

$$_{a}\bar{t}_{r} = {}_{m}D_{tr} + {}_{a}K_{tr} \quad ---- 1.$$

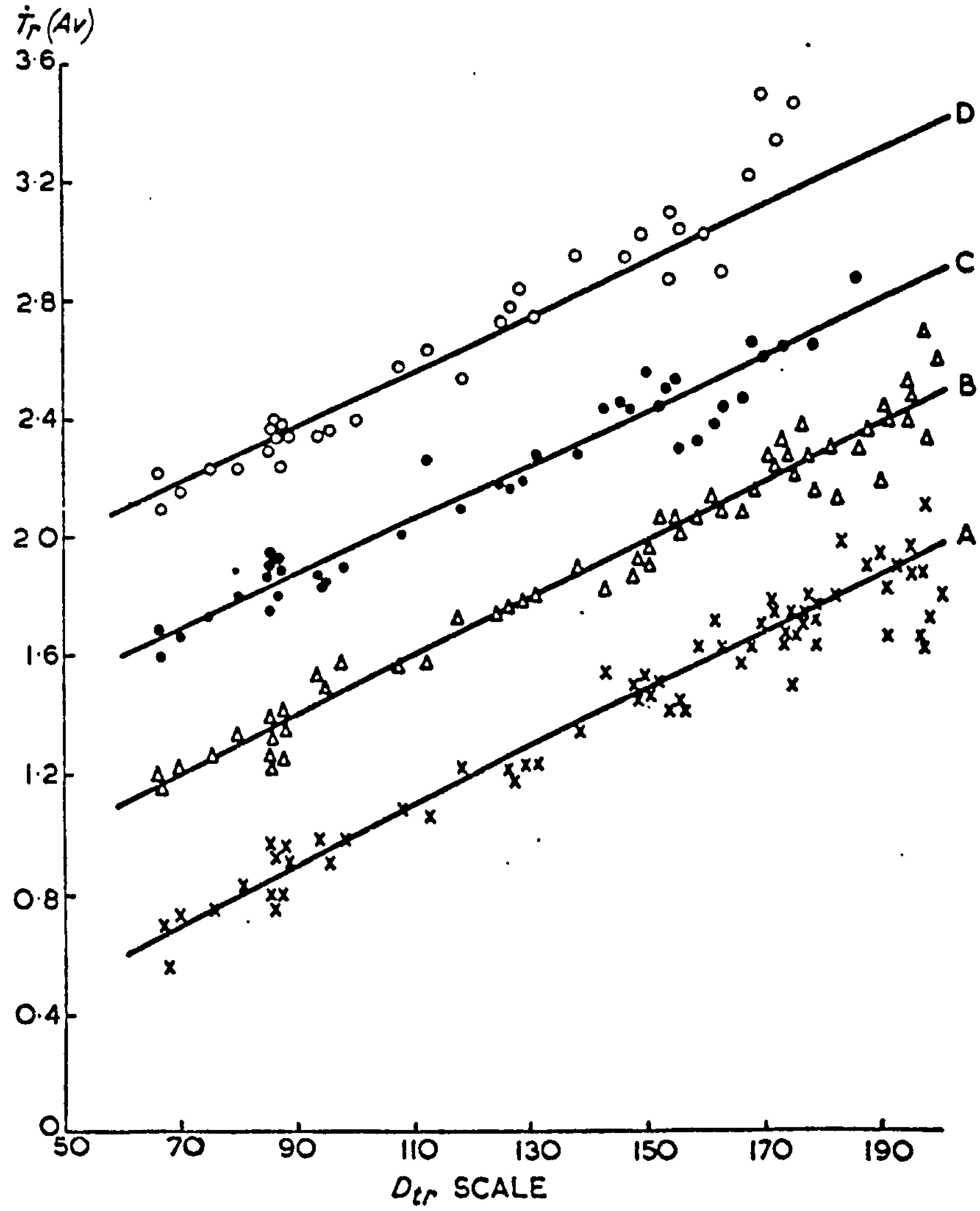


FIGURE 8. $\dot{T}_r(Av)$ PLOTTED AGAINST $100D_{tr}$ FOR GROUPS OF VARYING SPEED

Note. The curves *A*, *B*, *C*, *D* show the relationships to which the data should conform: they are *not* the best-fitting straight lines. To facilitate the examination of this graph 1·5 log units have been added to each value of $\dot{T}_r(Av)$ in Group *D*, and values of 1·0 and 0·5 log units in Groups *C* and *B* respectively.

Taken from Furneaux (1961)

where $_a\bar{t}_r$ = the mean item correct solution time for a subject '_a_' expressed in raw time units.

m = a slope coefficient

$_aK_{tr}$ = speed constant for the subject.

D_{tr} = item difficulty

On empirical testing, this function was found to be negatively accelerated. However, using log 10 as a transformation for the time measures (Tate 1950), the function achieved the linear form.

Equation 1 thus becomes

$$_a\bar{T}_r = {}_mD_{tr} + {}_aK_{tr} \quad --- \; 2.$$

where $\bar{T}_r = \log_{10} t_r$.

Following a series of intricate data manipulations (described in detail later), a point is reached at which the model can be subjected to a crucial test. The scaling computations generate a D_{tr} scale value for each item, and a speed constant and mean abandonment time for each subject. The test requires that equation 2 is suitable for all subjects, given the data produced by the scaling procedures.

This test is carried out by examining the relationship between the average correct solution time for subjects plotted against item difficulty. Furneaux's test first requires the subjects to be divided into 4 groups on the basis of their speed scores and then checking that the relationship between solution time and difficulty is similar in these sub-groups.

The data for this test are shown in Fig. 8, taken from Furneaux (1961).

In producing the data for Fig. 8, Furneaux divided his subjects into 4 groups, group A being the fastest, group D the slowest. It should be noted that in order to distinguish these groups graphically, a different constant was added to each group. Hence, the plotted group data are artificially separated. Details as to how this was accomplished are presented in the note accompanying the figure.

A second point to note about Fig. 8, is that the straight lines A to D are those to which the data _should_ conform. They are _not_, as Furneaux emphasised, the best fitting straight lines. $\dot{T}_r$ (Av) values are the mean log correct solution times for each problem averaged across subjects within each speed group. A description of the procedures for computing these values is presented in the chapter on the construction of D-scales.

Although Furneaux has provided details of the various statistical tests carried out on the data in nearly every instance in the course of D-scale construction, he did not do so when presenting his crucial test. The model requires that the functions be straight lines with unit slopes. No test of these requirements is reported and the lines A to D are, as noted above arbitrarily inserted. Furneaux has not published his raw data, so that an independant check on his conclusions cannot be undertaken.

Inspection of Fig. 8 reveals many more data points for group A than for other groups: that is, the faster the subjects, the greater the amount of data for them. Although this may seem unusual, it is a consequence of the procedures for computing $\dot{T}_r$ (Av) at higher difficulty levels. Only those items are included for which the data indicate that the items are below the critical difficulty level for the particular subject on that item: i.e. unambiguous inputs. This is related to the speed of the subject, as can be seen from Fig. 6. For very fast subjects, the lines AB would be shifted to the right, resulting in the region bounded by X and Z being shifted right. This means that the critical difficulty for that subject is higher and therefore more items are available which are not ambiguous.

Despite the absence of a test of the outcome presented in Fig. 8, the data do appear, on visual inspection, to conform to the requirements of unit slope in each of the four groups, satisfying the relationship expressed in equation 2.

Another of the requirements for equation 2 is a demonstration that variances between subjects and between items, at different levels of difficulty are homogeneous. Failure to achieve such homogeneity would severely restrict the applicability of this equation. Furneaux tested item and subject homogeneity by examing differences between subjects on unambiguous items and differences between items, on data derived from correct solutions which were unambiguous inputs.

In the process of carrying out these tests, two item pairs were found to produce inconsistent outcomes. When these items were deleted from the analysis, item and subject homogeneity of variance were found. The respective Chi-squares were 227 (p approx 0.4, 216 d.f.) and 36.1 (p approx 0.18, 29 d.f.)

On the basis of these analyses, Furneaux was able to conclude that the value of the log correct solution time variance "is independent of difficulty, and identical for all subjects". He does however point out that this solution was achieved only after deletion of data from two item pairs.

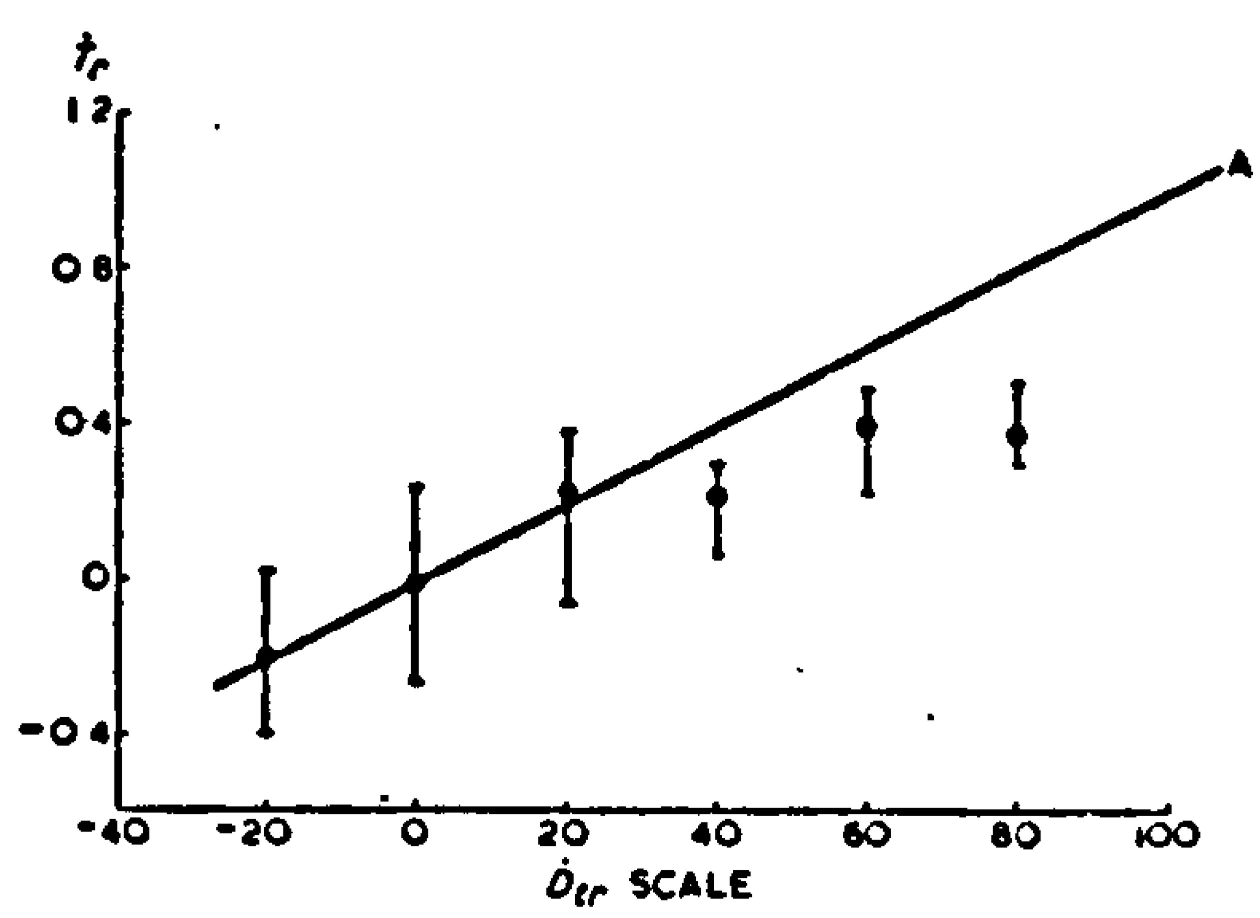


FIGURE 9. DISCREPANCY BETWEEN RELATIONSHIP OF $\dot{T}_r$ TO $\dot{D}$ WHEN I_a INPUTS ARE USED, AND THAT APPLYING TO I_u

The straight line A gives the relationship applying in the case of I_u inputs. Each point plotted shows the median observed value of $\dot{T}_r$ which relates to the set of inputs I_a having values of $\dot{D}_{tr}$ within the range $\pm$ 10 of the central value shown. The vertical lines terminate at the values of the upper and lower quartiles.

Values of $\dot{D}_{tr}$ have been multiplied by 100.

Taken from Furneaux (1961)

Furneaux has emphasised the importance of only using data from items that are unambiguous inputs. Items with D values above the critical level introduce the confounding effects of continuance which then make it inappropriate to combine or compare them with items not subjected to these influences. The model does not exclude the possibility of a subject passing an item beyond his critical difficulty level. However, do such items need to be differentiated from those correctly solved at difficulty levels below the critical value? (The former items are symbolized I_a (t_r) and the latter as I_u (t_r)). The question to be resolved from the data is whether or not a distinction is empirically important. That is, would relationships already uncovered, namely those of unit slope and of homogeneity of variance, still hold if such a distinction were not made?

Furneaux's findings on these questions are reproduced in Fig. 9, from which it can be seen that the data are neither linear in slope nor of equal variance. There are no statistical analyses reported to support these conclusions and the reader has again to rely on visual inspection.

On the basis of his study, Furneaux concludes that the data justify the retention of his hypotheses that 'm' in equation 2 and V "are population constants". In its final form, equation 2 is expressed as

$$T_r = m\ D_{tr} + K_{Tr} + E_i \quad - - - - \quad 3.$$

where E_i = a positive or negative component arising on a particular occasion i. This error component which is assumed to be distributed about a mean of 0 and variance V_{tr} over a large number of occasions.

Furneaux has commented on the unusual nature of his finding a constant in an area saturated with individual differences. The existence of a constant does not preclude individual differences, the latter being absorbed in individual differences in speed and continuance. He also suggests that equation 3 indicates a relationship analogous to Fechner's Law. The potential significance of this finding has been commented upon by Eysenck (1967a) and Butcher (1968), as indicated earlier.

This examination of the empirical findings of Furneaux's study has highlighted two broad groups of problems which make Furneaux's conclusions questionable. Firstly, the procedural details are inadequately reported, and of those published certain inadequacies remain. An instance of the former is the absence of the instructions given to his subjects; of the latter, his procedures for correcting item times for "apparatus time". The present writer has available all of the relevant reports (published and unpublished) referred to in the 1961 paper (Furneaux 1953, 1955) and none of these contain the necessary information.

The second broad group of problems are those relating to the absence of crucial statistical tests on the data. Undoubtedly, some relationships and group differences found in psychological research are such that statistical tests would be superfluous. It is also accepted that significance testing is only an element in the complex process of theory development and that the routine use of tests of significance for the sake of 'respectability' has been strongly critized (Bakan 1966; Lykken 1968). Nevertheless, certain statistical assessments would appear to be crucial in Furneaux's study. In particular, the relationship between speed and difficulty should at least have been assessed for higher order variance components (i.e. quadratic and cubic etc.) to see if a strong linear component emerged that accounted for most of the variance. It is all too easy to mask such higher order trend components if only graphical data presentation is used.

The questionable nature of Furneaux's findings is further highlighted by two other features of his procedures. The first of these will be considered in greater detail in a later chapter. At this point, it only need be noted that some of the item difficulty values had to be subjectively estimated in the course of scaling. Secondly, Furneaux was only able to achieve an adequate solution after deleting the data from two item pairs. These were reported to have distorted the variance of the solution time distribution and without their deletion, the analysis could not have proceeded.

In view of the above observations, and given the potential importance of Furneaux's findings, some form of replication is warranted: unless independent investigations are found to support Furneaux's claims, his findings must remain highly questionable, due to the inadequacies of his study described in this chapter.

Three forms of replication are possible, literal, operational and constructive (Lykken 1968). Literal replication is not possible, nor is it desirable given the inadequacies of group-testing data. The latter also makes it difficult to undertake a full operational replication, which requires the use of a similar sample and experimental procedures. Even if these constraints did not exist, constructive replication would still be necessary, given Furneaux's generalized claims. As Lykken states

> " - - - an experiment which replicates operationally but not constructively could be said to have demonstrated something, but not the relation between meaningful constructs generalizable to some broad reference population, which the author originally claimed to have established".

The present investigation was conceived in an attempt to test the generality of Furneaux's findings. It's specific concern was to investigate Furneaux's model in terms of it's applicability to a different set of test items using a different procedure for administration and a sample which also differed from that used by Furneaux.

3. Furneaux's 'Problem-solving Mechanism'

In addition to presenting a conceptual analysis of problem-solving performance, and a procedure to scale test items, Furneaux attempted to develop within the context of neuropsychology a 'problem-solving mechanism'. The goals of this model were to encompass existing data as well as to generate testable predictions about its functioning. The mechanism eventually proposed was based on findings reported by Hick (1950).

Hick (1950) was able to show that reaction-time in a multiple choice situation was related to the complexity of the task. This relationship takes the form

$$RT = K \log M$$

where RT = choice reaction-time
K = an individual constant
M = a function of the complexity of the choice situation.

Hick (1950) argued that if multiple choice activity involved successive binary classifications, a relationship of the type indicated in the above equation would be anticipated. The structural model proposed by Hick (1950) is in some ways analogous to a computer memory. He suggests that as the brain is composed of seemingly identical elements (cells), all its activities will involve some form of binary switching, an operation which will involve sequences of components similar in kind and duration. These sequential switching operations take the same basic time and involve the same activity. Furneaux's interpretation of Hick's (1950) hypothesis is that it implies that the multiple-choice reaction time provides a measure of the "time required for a search to be completed in the brain for the set of "connexions" which would initiate a required behaviour". Furneaux's development of this hypothesis was to suggest that problem-solving could be regarded as a special case of multiple-choice reaction-time.

The hypothesis eventually devised by Furneaux proposed an elaborate mechanism entailing a number of assumptions and restrictions, which, in the final analysis he says should not be taken too seriously.

The basic structure of this model is a set of elements which are brought into some "interconnexion" when a problem is presented. A randomly selected set is initially brought into operation and tested for correctness by a "comparator". If the comparator accepts the solution, it is output as a response. If the solution is rejected, the search process is re-instigated. A more complex set of elements is then brought into operation and the process continued. As such processes take time, and as the time taken is a function of the complexity of the problem, one can readily see how this model relates to Hick's (1950) equation. In Furneaux's model, continuance is accounted for by the comparator being set to operate for some period when the search process is instigated.

Although Furneaux has tested a number of predictions from this model, and while only some of these have received support, there appears to have been no further development of his ideas. However, in 1964, Roth published a paper in which he reported a significant correlation between a rate of gain measure from a choice reaction time task and a measure of intelligence. On the basis of this finding, it appeared that Furneaux's problem-solving mechanism might merit greater attention. Furneaux had not incorporated a choice reaction time study in his research and was therefore unable to make any comparisons between his M-statistics and any of the indices derived from the choice situation. The present study did however present an opportunity to do so. Hence, the second research problem investigated here concerned the relationship between reaction-time and problem solving speed.

4. Further Aspects of Furneaux's Research:

Most of the detailed information presented in Furneaux's 1961 paper is confined to the investigation of correct solution times and their associated D-scales and M-statistics. Although he has investigated incorrect solution time and accuracy data only the outcomes of these investigations are reported. This chapter will present a summary of these investigations.

In general, Furneaux's strategy in these further analys~~i~~es followed the same ~~interactive~~ iterative procedure devised for the Dtr scale construction but with some modifications to specific components of the procedure. Furneaux does not specify the changes which he introduced and in the absence of the relevant data, his findings are difficult to evaluate.

Furneaux reports finding that the mean times to correct and incorrect solutions below the relevant critical difficulty levels differed, with the incorrect time mean being the smaller. With increasing D values, the differences between means increased in magnitude. The variance of the incorrect solution times was greater than that found for correct solutions, and the disparity between them increased with increases in difficulty.

The incorrect time data however were found to be particularly subject to individual differences and no simple functional relationships could be established. Furneaux noted that incorrect times tend to show poor reliability, partly as a consequence of their wide variance, and because there were very few incorrect responses at low difficulty levels, insufficient data were available to allow reliable estimates to be obtained.

In his particular groups of subjects, the correlation between correct and incorrect times was 0.58 at low levels of difficulty, and fell to below 0.5 at higher difficulty levels. The important conclusion derived from these comparisons was that correct and incorrect times are not interchangeable. Thus any measure (for example of rate), which combines such times is likely to be an agglomerate score unsuited to precise investigation.

Furneaux reports having had more success in the analysis of accuracy data. He found that using the already established D_{tr} scale values, and with K_{qr} values computed, it was possible to "specify a normal ogive which - - - gives an acceptable fit to the data relating q_r to D_{tr} - - -". The D_{qr} scale finally evolved correlated 0.92 with the D_{tr} scale. This finding led Furneaux to conclude that, for the tests used, the determinants of " response-time and response-quality, are much the same". Thus, it is legitimate to call both "difficulty" scales. It should be noted here that it is only D-scales for speed and accuracy that are correlated. The fact of their correlation does not allow a generalization that the individual constants K_{Tr} and K_{qr} are correlated to the same extent. The evidence on this point is presented later.

On the basis of the data presented it is possible extract what in the final analysis, is a simple but often ignored principle. This principle has influenced all of Furneaux's work. It states that any combination of data must be preceded by empirical evidence that such a combination is permissable. A logical analysis into irreducible components is a necessary but not sufficient basis for proceeding with data combination. Empirical evidence from a variety of sources is necessary before such combinations can take place. It is the failure to do so which tends to invalidate much of the current and early work on human problem solving.

Further illustrations of the inadequacies of agglomerate scores are presented by Furneaux. He examined the consequences of using wrong solutions and abandonments as indices of "incorrect" performance when the goal of the analysis was to obtain the proportion of incorrect solutions in a particular group. Such a procedure, as he points out, leads to a multidimensional scale whereas the presumed intention was a scale with

TABLE 8. Intercorrelations of K_{Tr}, Kqr and $\bar{T}s$.

Decimal point omitted.

	K_{Tr}	$\bar{T}s$
Kqr	-38	31
$\bar{T}s$	-27	

NOTES.

a. A high value of K_{Tr} implies slow speed

b. A high value of Kqr implies high accuracy

c. A high value of $\bar{T}s$ implies high continuance.

(Data based on 209 subjects, a sub-group of the research sample. It was not possible to get all three indices from all subjects.)

Table taken from Furneaux (1961, Table V.1) unidimensional characteristics. For slow subjects, such a D-scale correlated only 0.68 with a previously established D_{tr} scale whereas for fast subjects, the corresponding correlation rose to 0.79.

Even under conditions where difficulty values are closely correlated, as in the already mentioned accuracy/correct solution D-scales, one cannot assume that M-statistics for the same data are closely related. Evidence on this point is contained in Table 8.

It will be obvious from the data in Table 8 that there is a substantial degree of independence among these scores. Furneaux has found evidence that within a more homogeneous group the interrelationships are even weaker.

In view of the aforementioned evidence, as Furneaux points out, it would be inappropriate to conceptualize such abilities as "letter-series test ability". One can extend this view and question the meaningfulness of other conceptions of ability which presuppose a homogeneity among its various indices. The more appropriate conception would be in terms of speed, accuracy or continuance within a particular problem area. The theoretical and practical implications of this conclusion are profound for it means that a fundamental reorientation in studies of human abilities is required. Instead of intercorrelating agglomerate scores, the researcher needs to seek out relationships using homogeneous indices, rather than retaining the traditional crude approach.

In his discussion of these conclusions, Furneaux raises the question of evidence for a "speed factor" in problem solving. Would speed constants derived from different types of test material intercorrelate to the extent that the presence of such a "factor" can be presumed to 'exist'? No studies known to the present writer have attempted to answer this question using Furneaux's measure of speed. Hence, one of the tasks of the research reported in this study is that of examining the interrelationship among measure of speed derived from two different types of test material to see if they show a substantial intercorrelation. It is only from studies of this type that an appropriately measured speed factor can be derived.

Furneaux has examined the research literature for evidence of general factors of speed, accuracy and continuance. Although he gives insufficient details of the evidence examined he is led to conclude

> "There would thus appear to be quite a strong justification for talking of a subject's <u>speed</u> and <u>accuracy</u> without making specific reference to the kind of problem-material in terms of which these attributes were measured" (P.188)

This view is open to question. Firstly, Furneaux's review of the evidence is limited. He cites only two studies (Rimoldi 1951; Thurstone 1949) which support the generality, across tests, of 'speediness'. Neither of these used his K_{tr} index, and, as he points out, the score or more other studies of speed (of which he provides no details), used inappropriate indices of speed. The measure of accuracy, K_{qr}, drived from the letter series test "usually displays correlations with cognitive test scores of the conventional kind which are higher than those associated with either speed (K_{Tr}) or continuance (C)". However, Furneaux does not present any evidence in support of this statement. While he cautions that this relationship is not invariant, he does suggest that since conventional scores tend to be quite highly intercorrelated, and since his accuracy measure is also correlated with such scores, it is concluded that there may be quite a strong factor of accuracy involved in a "wide variety of cognitive performances". Such evidence, in the view of the present writer, is inadequate as a basis for the broad generalizations made by Furneaux.

The second set of considerations which must be taken into account before drawing conclusions as to the generality (across tests) of speed and accuracy are those concerned with the effects of instructions, and orectic factors on test performance.

Furneaux is much less certain about the generality of continuance. In the case of this variable, he implicates the instructions given to subjects. If the instructions are presented in a particular way, persistence and continuance become synonymous. Under these conditions, there is some, again limited, evidence for one or two related traits. However, the relationships are so tenuous that Furneaux at least would not be surprised if continuance were found to be test specific.

The concepts of <u>power</u> and <u>level</u> refer to attributes which, as Furneaux points out, are not yet defined in any generally acceptable way. His own use of the term <u>level</u> is reserved for tests designed in such a way as to maximize persistence. Furneaux (1953) has devised a series of tests of level, constructed as a series of cycles each containing items of increasing difficulty, the cycles themselves increasing in overall difficulty as the test progresses. The subject is allowed unlimited time and is encouraged to persist when faced with difficult items. The items themselves have not been scaled according to Furneaux's own procedures for D-scaling (Brierley 1969) and as he explicitly recognises, the scores produced by these tests are subject to many of the limitations associated with conventional tests. Thus, he regards such scores as only an approximate measure of the level of difficulty at which the testee can function, given encouragement to persist with items experienced as difficult. While he suggests that such tests "measure a reasonably invariant property of the individual", Furneaux does not present any evidence to support this assertion.

As has already been noted, Furneaux recognizes the potential effects of instructions on test performance, mediated by continuance. He also emphasizes the effects of orectic influences and reports one study which he has carried out to assess the effects of drive on speed. The index of speed in this study was the difference in score obtained when subjects were tested under stressed and unstressed conditions. In addition to their speed scores, Furneaux also obtained measures of extraversion and neuroticism using Guilfords S.T.D.C.R. inventory. His group of 75 students was divided into four sub-groups, depending on their personality test scores. These sub-groups were the stable introverts and extraverts, and neurotic introverts and extraverts. He found that these sub-groups reacted differentially to stressed testing and on these, as well as other grounds, implicated orectic factors in test performance.

5. A Summary of the Aims of the Present Research:

Furneaux's (1961) paper is replete with testable assertions. However, the most fundamental of these is the replicability of his scaling procedures with different tests, subjects and forms of item presentation. Thus the primary aim of this study was to undertake a scaling exercise on standard test items to see if the outcome was consistent with that reported by Furneaux.

In the course of difficulty scaling, various constants become available for each subject. These include speed, accuracy and continuance indices. It was the intention of the present study to test a number of hypotheses relating these data to Eysenck's (1967a) predictions about systematic personality differences in test performance. These hypotheses are detailed later.

A third aim of this study was to examine slope measures from a choice reaction time task in relation to the Furneaux speed scores and to measures of intelligence on the grounds that a number of hypotheses about such relationships can be derived from the theoretical model proposed by Furneaux (the problem solving mechanism) and the assertions made by Eysenck (1967a) regarding the significance of Roth's (1964) study.

VII. AN INVESTIGATION OF FURNEAUX'S DIFFICULTY SCALING PROCEDURES

1. Apparatus, Procedures and Subjects:

A LINC-8 computer (Digital Equipment) was used to control item presentation and to record item solution and solution latency. The control programme was written in LAP-6, a language specific to the LINC and designed specifically for on-line control and processing. At the time this study was undertaken, the LAP-6 compiler had been well tested in the U.S.A., as had the machine. Since then, both the compiler and the hardware have undergone a number of developments and modifications. For this reason, neither the programme nor the circuitry linking test administration apparatus to the computer are given in detail.

The control programme was specially written by the author and included a timing sub-routine for timing responses. This was necessary because at that time, there was no internal clock in the LINC-8. The calibration of the programme clock was checked against a millisecond timer. All time data used in the subsequent analysis were adjusted to the nearest 1/100 secs. It should be noted here that every programme specifically written for this study was thoroughly checked by the author before it was used. For example, the test administration programme was checked several times by the input of preselected responses and latencies.

a. Tests

i) Personality Measures:

The P.E.N. (Psychoticism, Extraversion and Neuroticism) Scale was used to provide P, E, N and L (Lie-Scale) scores for the subjects. This scale is composed of E, N, and L items from a pool of items developed by the Esysencks (Eysenck and Eysenck 1968) and the norms were obtained from one of the authors (S.B.G. Eysenck). At the time of this study, the P items had been selected following an extensive set of studies carried out with various normal and psychiatric groups (Eysenck and Eysenck 1968).

A copy of the test is incorporated in the Appendix.

ii) Intelligence Tests:

The tests chosen for scaling were the multiple choice form of the Mill Hill Vocabulary Scale, Form 1 Senior, Set B (Raven 1965b), and the Advanced Progressive Matrices, Sets I and II (Raven 1965a). In addition, items from the Standard Progressive Matrices (Raven 1960) were used. Details of the test are presented below.

FIGURE 10. Conventional difficulty of Matrices cycles.
Row 1 = Test order
Row 2 = Matrices number.

TABLE 9. The 38 items of the cyclical form of the Matrices.

PRELIMINARY ITEMS

TEST ORDER	SOURCE	ORIGINAL NUMBER
1	A	5
2	A	6
3	B	1
4	A	8
5	B	3
6	A	10
7	B	5
8	A	12

CONCLUDING ITEMS

TEST ORDER	SOURCE	ORIGINAL NUMBER
33	A	7
34	B	2
35	A	9
36	B	4
37	A	11
38	B	6

CYCLE

TEST ORDER	SOURCE	ORIGINAL NUMBER	TEST ORDER	SOURCE	ORIGINAL NUMBER
	1			2	
9	B	7	13	B	8
10	C	1	14	C	2
11	C	14	15	C	17
12	C	30	16	C	31
	CYCLE				
	3			4	
17	B	9	21	B	10
18	C	5	22	C	6
19	C	16	23	C	19
20	C	33	24	C	34
	CYCLE				
	5			6	
25	B	11	29	B	12
26	C	7	30	C	8
27	C	20	31	C	21
28	C	35	32	C	36

KEY:
A = Standard Progressive Matrices
B = Advanced " " Part I
C = " " " Part II

There were several reason for selecting these particular tests. In the first instance, they were of a conventional difficulty likely to discriminate among the subjects available for this study. Secondly, they were sufficiently different from the letter series used by Furneaux to justify their being regarded as appropriate for a constructive replication. Thirdly, both are well established tests, in wide use. Fourthly, they appear to tap somewhat different types of ability. For example, Matrices (as well as Furneaux's Letter Series) load strongly on the fluid intelligence factor whereas vocabulary loads more strongly on the factor identified as crystallized ability (Horn 1970, 1972; Cattell 1971). These two tests also have loadings on different factors in the heirarchial structure model, Matrices loading predominantly on 'g' and vocabulary tests on 'v : ed' (Vernon 1961). Superficially, the two tests also appear to require somewhat different skills. Whereas the Mill Hill items appear to be such that the subject either knows or doesn't know the answer, the Matrices require a solution to be 'worked out'. Finally, the Matrices are published with details of the conventional difficulty of the items, so that it is possible to devise 'cycles' of difficulty without first having to pilot the items on a comparable sample.

The Mill Hill items were administered in their original order. Items from the Standard and Advanced Matrices were rearranged so as to form 6 cycles of 4 items each. Within each cycle, items were arranged in ascending order of conventional difficulty. Each cycle in turn increased in difficulty. The cycles were preceeded by 8 items from the easiest parts of both sets of Matrices and followed by 6 items also of low difficulty.

This cyclical arrangement was adopted by Furneaux (1955) in constructing his level test. The purpose of such mixing was to encourage persistence although it has never been demonstrated that this arrangement achieves its goal. The arrangement of the Matrices items is presented in Table 9. The conventional difficulties are shown in Fig. 10, with presentation order and original Matrices item number in the row beneath.

In all, there were 33 items in the Mill Hill and 38 items in the cyclical Advanced Matrices.

It should be noted that the first item of the Mill Hill was used as an example. Hence the version used here contained one item less than the standard test.

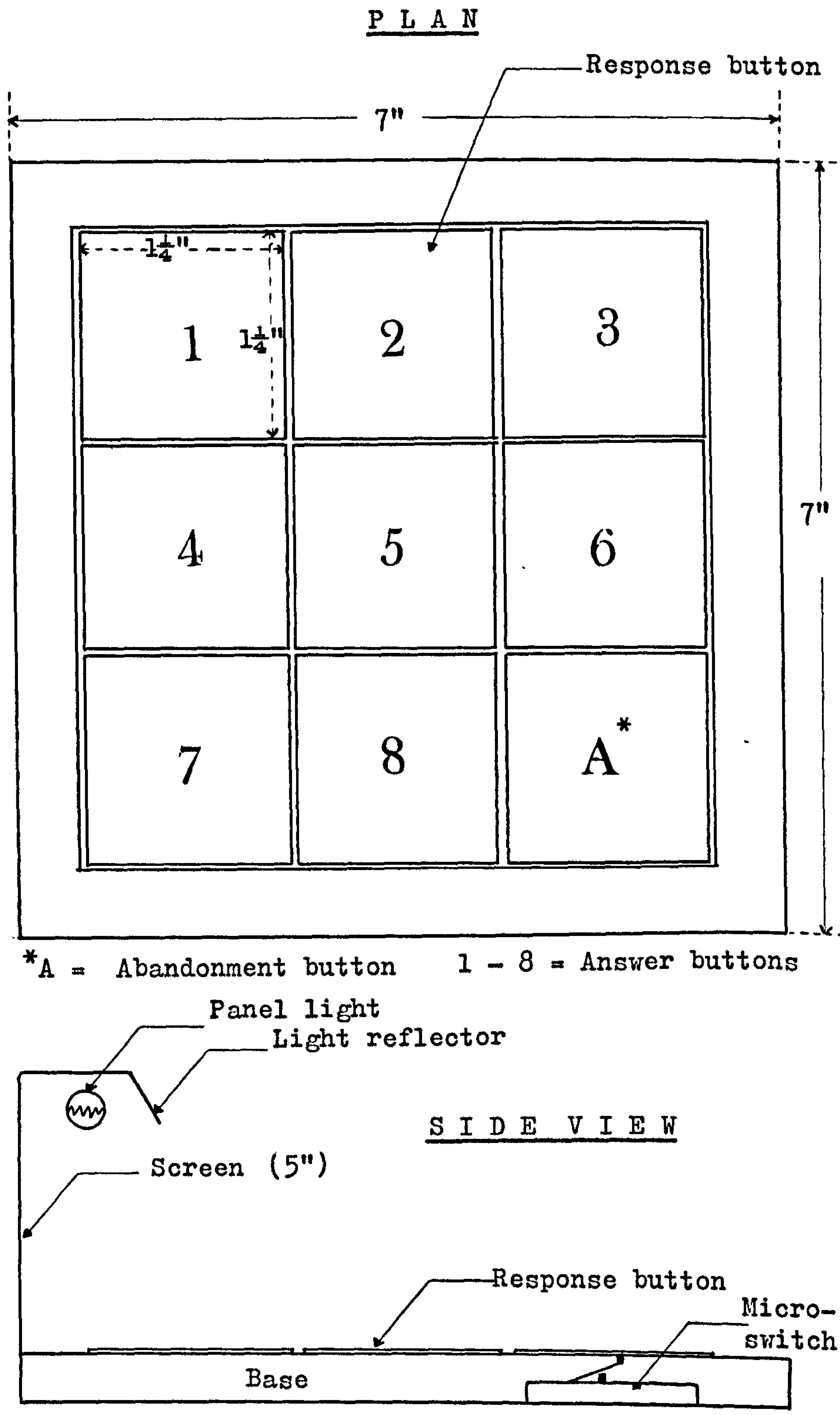
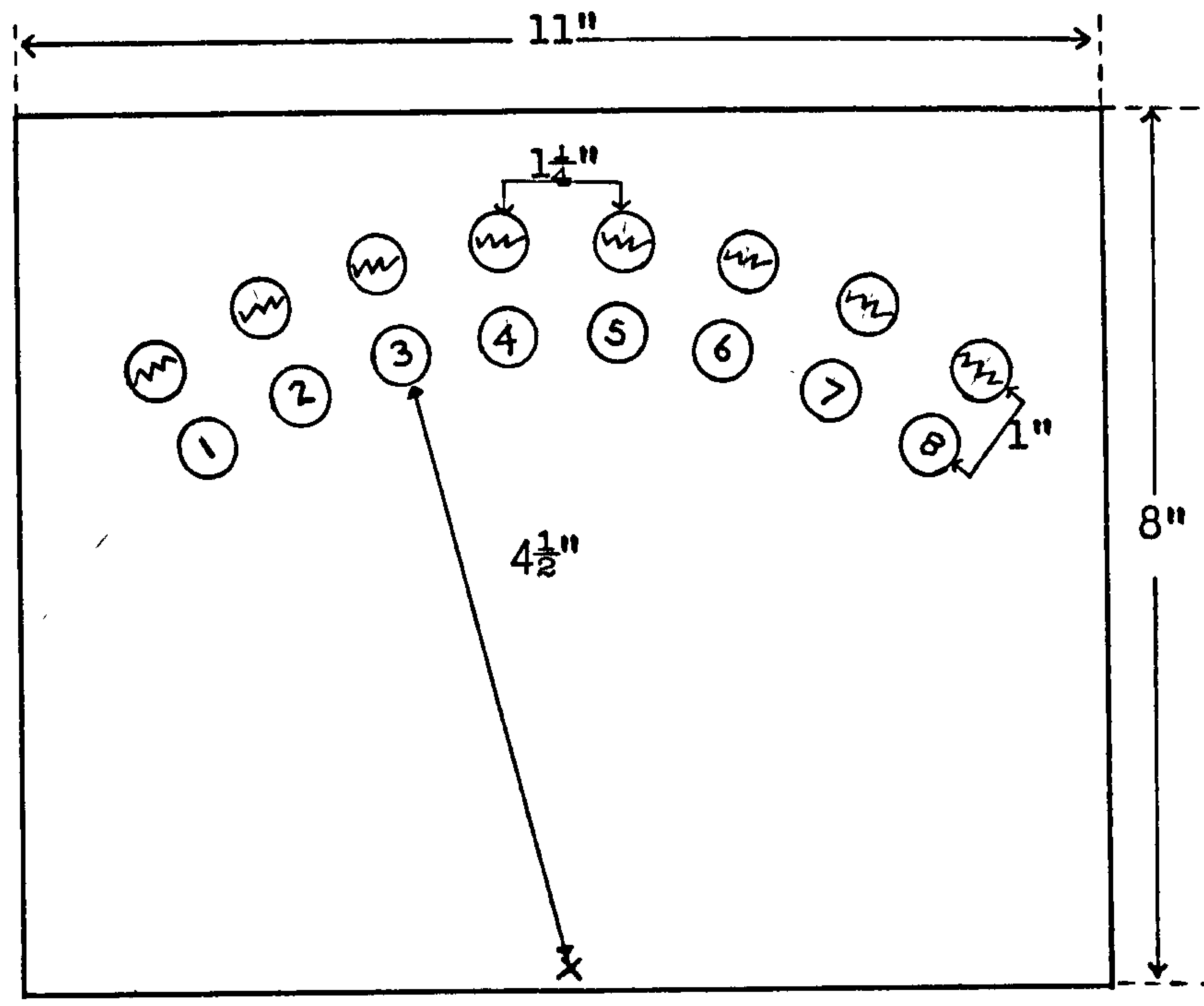


FIGURE 11. Multiple-choice Response Apparatus

P L A N



KEY
Pea Lamp, 2.8 Watts, Green dome.

Response button, Red.

× Starting position.

S I D E V I E W

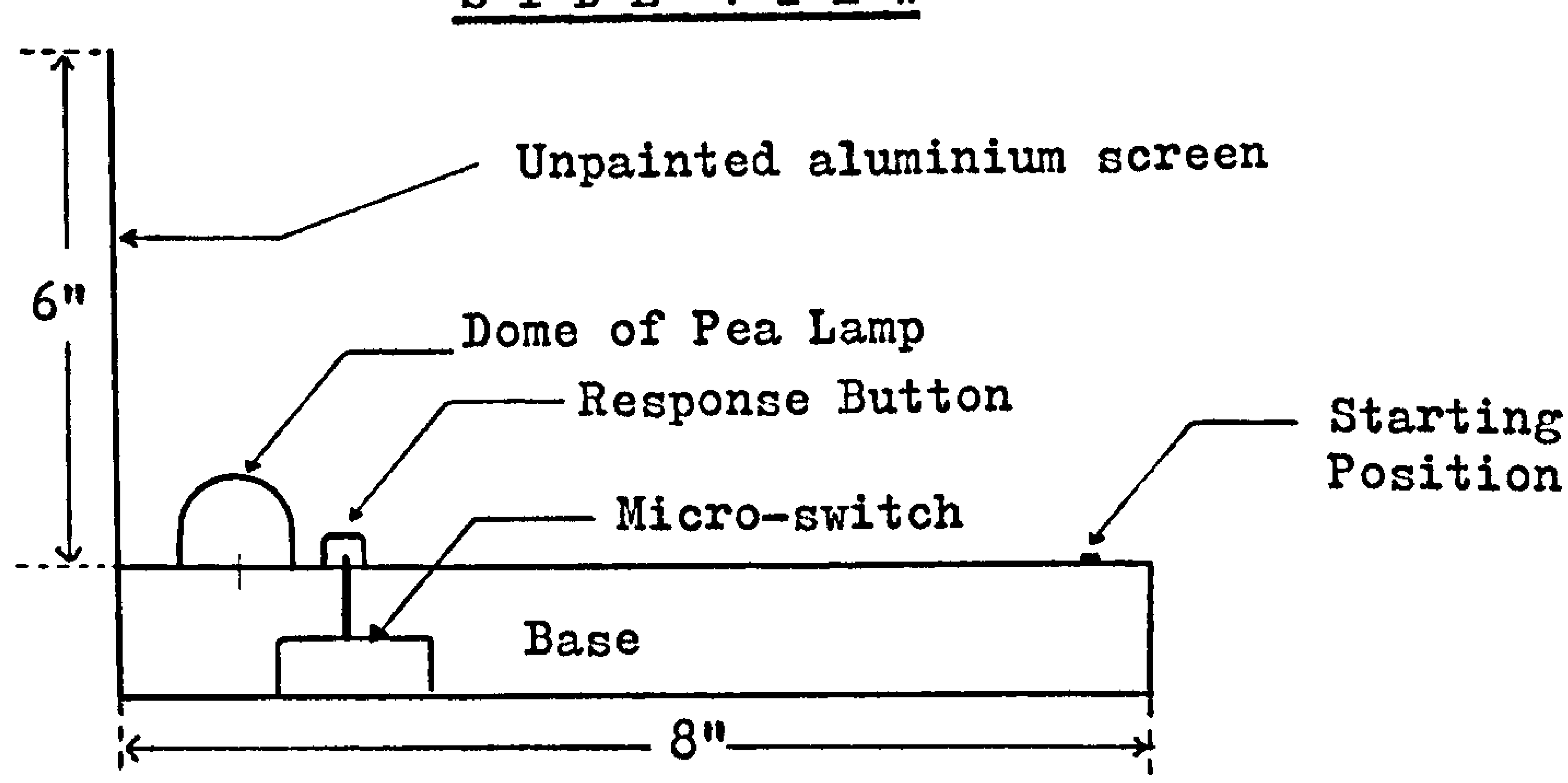


FIGURE 12. Choice Reaction Time Apparatus.

The test items were photographed using positive black-and-white 35 m.m. film and transferred to individual slides. The slide magazine was positioned so that subjects could not readily discover which was the last item in the series. This precluded the possibility of an 'end-spurt'. This form of presentation also prevents subjects looking ahead and does not allow them to return to any item once it has been completed. These factors are difficult to control in group testing with standard paper and pencil tests.

A major advantage of the procedures adopted in this study is that they enable surreptitious recording of solution times.

Test items were presented by back-projection onto a ground glass screen using a Kodak Carousel 35 m.m. projector. All items were of the multiple-choice type. The subject was required to press one of 9 keys to indicate his solution. One of the keys was a "Don't Know" key used to record an abandonment. On completion of testing, the computer simultaneously output a teleprinter record of solution and solution latency and a punched tape. These tapes were subsequently used to provide a set of 80 - column data cards used in later data analysis. All such analyses were carried out on the University of London C D C 6600 computer.

All peripheral equipment was constructed in the Psychology Department workshop at the Institute of Psychiatry. The apparatus and its dimensions are detailed in Fig. 11.

iii) Choice Reaction Time:

The presentation of the reaction time stimuli was controlled by the LINC-8 computer.

The design of the apparatus followed the specifications given by Roth (1964). It consisted of a semicircular arrangement of 8 pea lamps, all covered by green translucent perspex. Immediately in front of each lamp was a response button, 1 inch from each lamp. The lamps and buttons were mounted on a board which had a starting position so located that it was equidistant from the response buttons. Subjects were required to rest the index finger of the preferred hand on this point, and to respond by moving from this point to depress one of the response keys, returning to the central point at the end of each reaction. The response buttons were directly linked to micro-switches. The response board and its dimensions are shown in Fig. 12.

The reaction recorded was the preferred hand response time between the onset of one of the lights and the pressing of its corresponding microswitch. For the simple reaction time series (1/1), subjects were required to press the switch immediately to the left of the centre of the array:for 1 of 2 choices (1/2), either of the two middle buttons had to be pressed; for 1 of 4 (1/4), the four central buttons; for 1 of 8 (1/8), any of the 8 buttons. When buttons were not required for a particular series, they were covered by special masks which covered both the pea lights and the response keys.

The onset of the stimulus lights was controlled through the LINC-8. This programme also recorded response latency and which of the buttons was pressed. The lights were switched on for a maximum of 1 second. If no response was made before this, the light was switched off automatically and a response code of 9 recorded for that stimulus. These times were not included in any of the calculations.

Each reaction time set was made up of 40 stimulus presentations. Each subject was given 9 sets of reaction time trails, arranged as follows:-

1.	Practice	5.	1/8	8.	1/4
2.	1/2	(Mill Hill/Matrices)		9.	1/2
3.	1/4	6.	1/8		
4.	1/1	7.	1/1		

A rest period was allowed between each set and each set was repeated following the administration of the Mill Hill and Matrices Tests. In the second series, the order of presentation was the reverse of the order of the first. However, the sequence of lights and inter-stimulus intervals were identical in both series.

In the practice set, devised to familiarise subjects with the task requirements, the lights were presented in sequence, from left to right, 1-8, 8-1, 1-8 etc. until 40 practice trials were completed.

For the experimental series, the set of 40 trials was divided into 5 blocks of 8 trials per block, although the subjects were not aware of this structure and no gap existed between each block other than the normal interstimulus interval. The sequence of lights and inter-stimulus intervals in blocks 1, 3, 5 of each set was identical but determined on a random basis. Blocks 2 and 4 also contained random sequences of lights and inter-stimulus intervals that differed from each other as well as from the sequences in blocks 1, 3, 5. This psuedo-random series was included for purposes of a future investigation into performance variation over time.

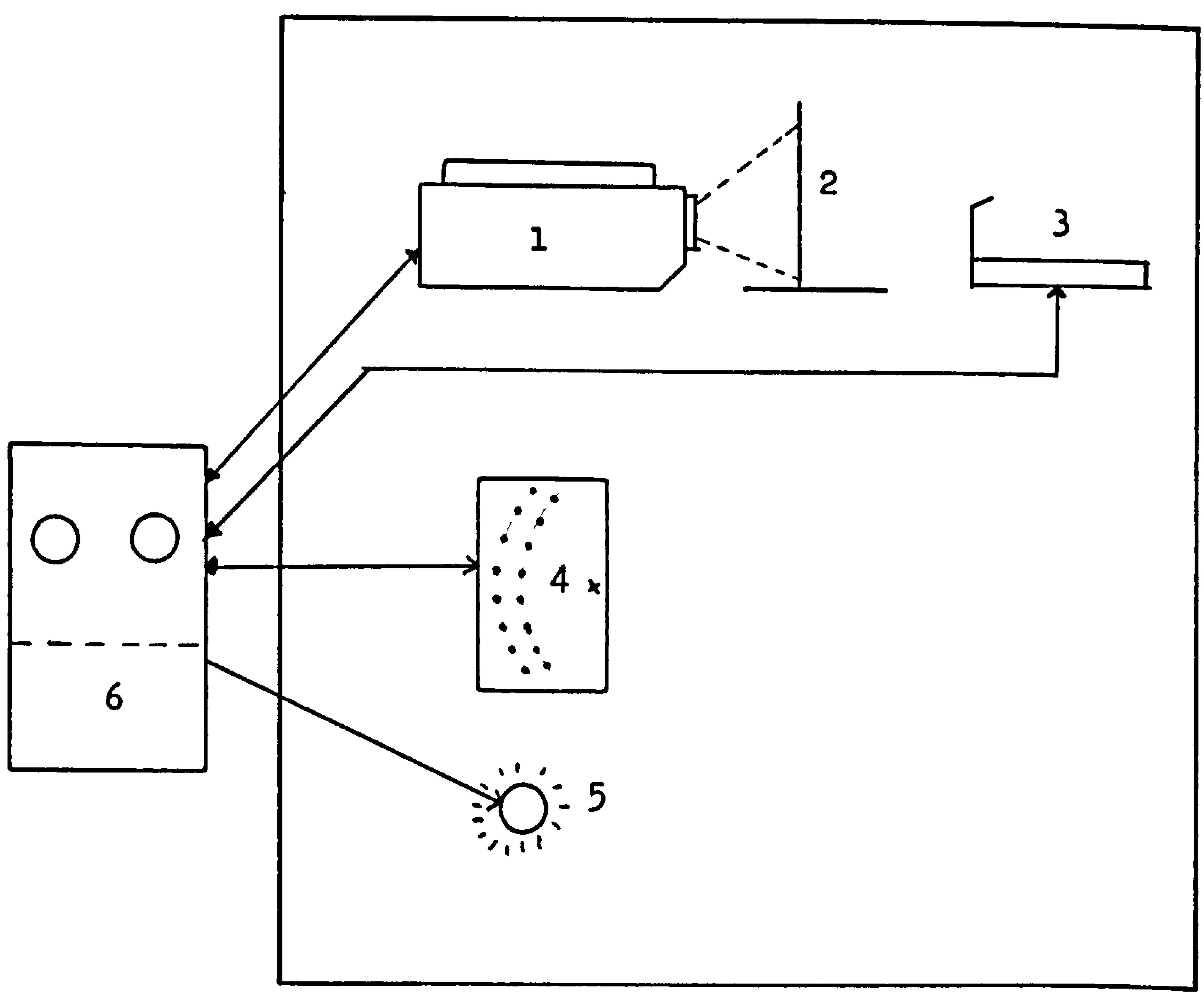


FIGURE 13. Schematic layout of apparatus and computer.
1=Slide projector 2=Screen 3= 9-choice panel
4=Reaction time lights and response buttons
5='Ready' light 6=LINC-8 computer

The only other constraint was that within each block, each stimulus should appear with equal frequency. The stimulus intervals varied between 2, 3, 4 and 5 seconds and their order was identical in each set. The order of stimulus presentation and the associated inter-stimulus intervals are given in the Appendix.

The times were rounded to the nearest 1/1000 sec.

b. Setting:

Testing was carried out individually. The subject was seated in front of the screen and response key panel in a darkened sound-proof cubicle. The panel was illuminated by a focussed light built into the panel. The computer was located in an adjacent room and there was no transmission of sound between these rooms. However, a one-way screen permitted the subject to be observed and a special signal light was used to indicate that the experiment was still in progress. This light was controlled by the computer and acted as a signal to the experimenter that one section of the study was ended and that he could enter the subject's cubicle to instruct him on the next section of the study. Fig. 13 is a schematic representation of the apparatus layout.

c. Procedures:

The following sequence of testing was used for all subjects:

P.E.N. (completed in a different room)
Reaction Time - part 1
Rest
Mill Hill
Rest
Matrices
Rest
Reaction Time - part 2

d. Instructions:

The full instructions are given in the Appendix.

In line with the aims of this study, instructions were designed to foster the measurement of "natural" speed on the Mill Hill and Matrices.

e. Subjects:

The subjects who participated in the present study were paid volunteers who came to the Psychology Department, Institute of Psychiatry to participate in this and another research project. The other project was an investigation of the effects of unconditioned stimulus intensity on the conditioned eyelid response (Evans 1970). There was no overlap between the two studies apart from the common use of personality data and other data on subject characteristics (age). As far as is known participation in one could in no way interfere with the outcome of the other.

All subjects were males and came from one of two larger groups. One group consisted of 'white collar workers', some of whom had responded to an advertisement in a weekly newspaper. Other members of this group were subjects known to the department who had previously participated in other projects. None of them had previously completed the Mill Hill or Advanced Matrices. The other group were students at a variety of institutions of higher education. None of these were psychology students, and so far as could be ascertained, had not been tested on the instruments used in this study.

It is obvious that the study sample cannot in any way be said to be representative of any group in particular. Given their occupational status, they were likely to be of at least average ability in standard tests of intelligence but probably of above average ability. Being volunteers, they also constituted a group with certain undetermined biases. These factors no doubt restrict the generalizability of any conclusions to be drawn from the results. However, as the main concern was with the scalability of test items, and as Furneaux regards the speed/difficulty relationship as population independent, the sampling constraints are not likely to be a major factor in the scaling given this 'normal group'. There is an important constraint on sampling imposed by the apparatus. The use of computer-controlled testing requires subjects to be brought to the apparatus. The machine is not portable. This is a constraint which was particularly important when this study was undertaken. With the advent of remote access terminals, such constraints are likely to be of lesser significance.

119 subjects were tested. This group was reduced to 110 for reasons given below.

Subject No.	Reason for exclusion
7	P.E.N. not given
8	P.E.N. incomplete
21	" "
24	Anxious, partially deaf, reaction times lost due to failure to respond.
52	Apparatus fault
55	" "

TABLE 10. Basic data for the 110 male subjects.

VARIABLE	MEAN	S.D.	RANGE	NORMS* MEAN	NORMS* S.D.
Age	29.16	10.67	18-65	-	-
P	1.86	1.68	0-7	2.50	2.71
E	11.28	4.22	2-20	12.75	4.12
N	9.00	4.66	0-19	7.33	4.37
L	2.93	2.27	0-10	-	-

* Normative data provided by Dr. S.B.G. Eysenck.
Lie scale (dissimulation) data not available.
Male norms only.

FIGURES 14-17:

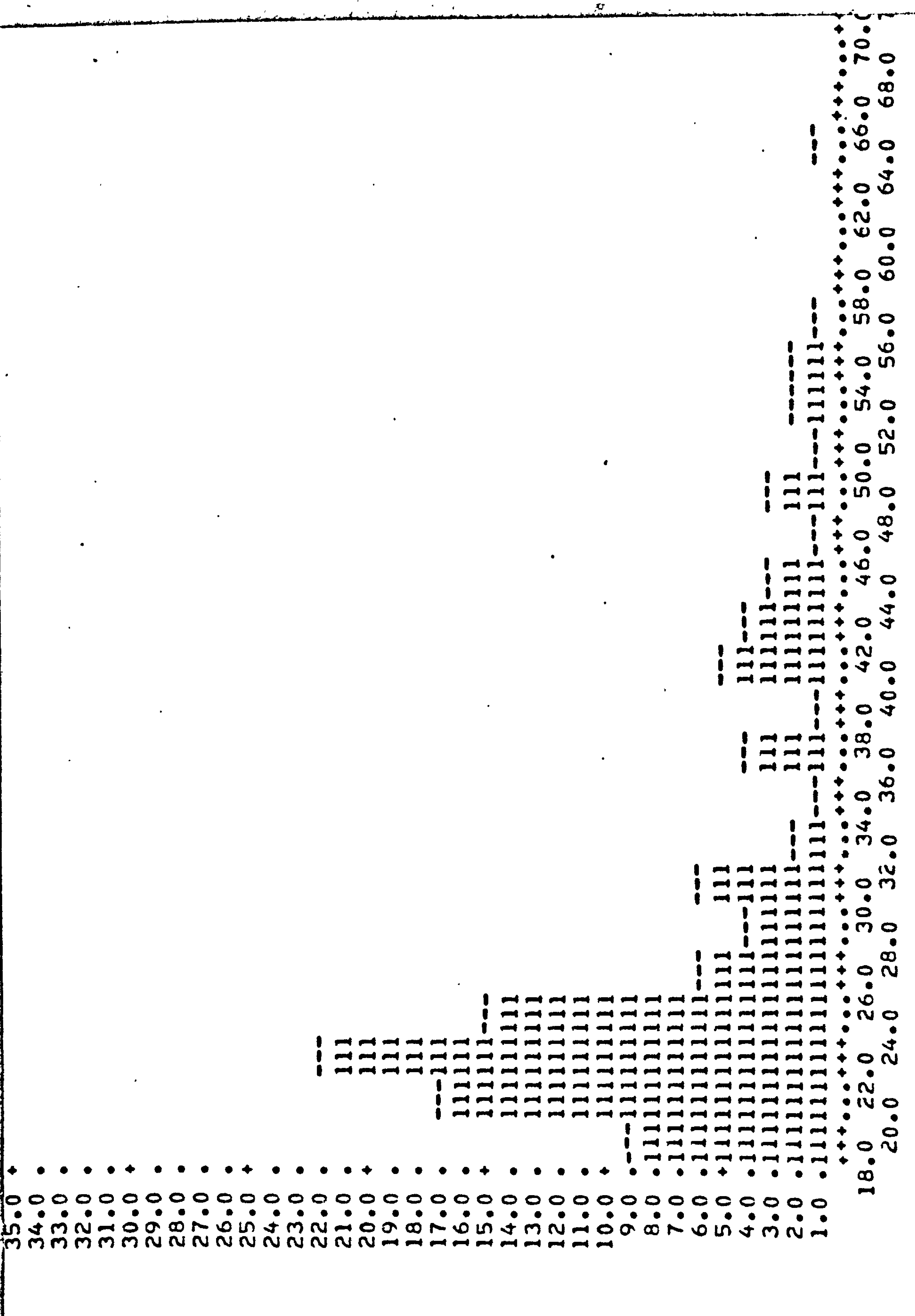


FIG. 14.

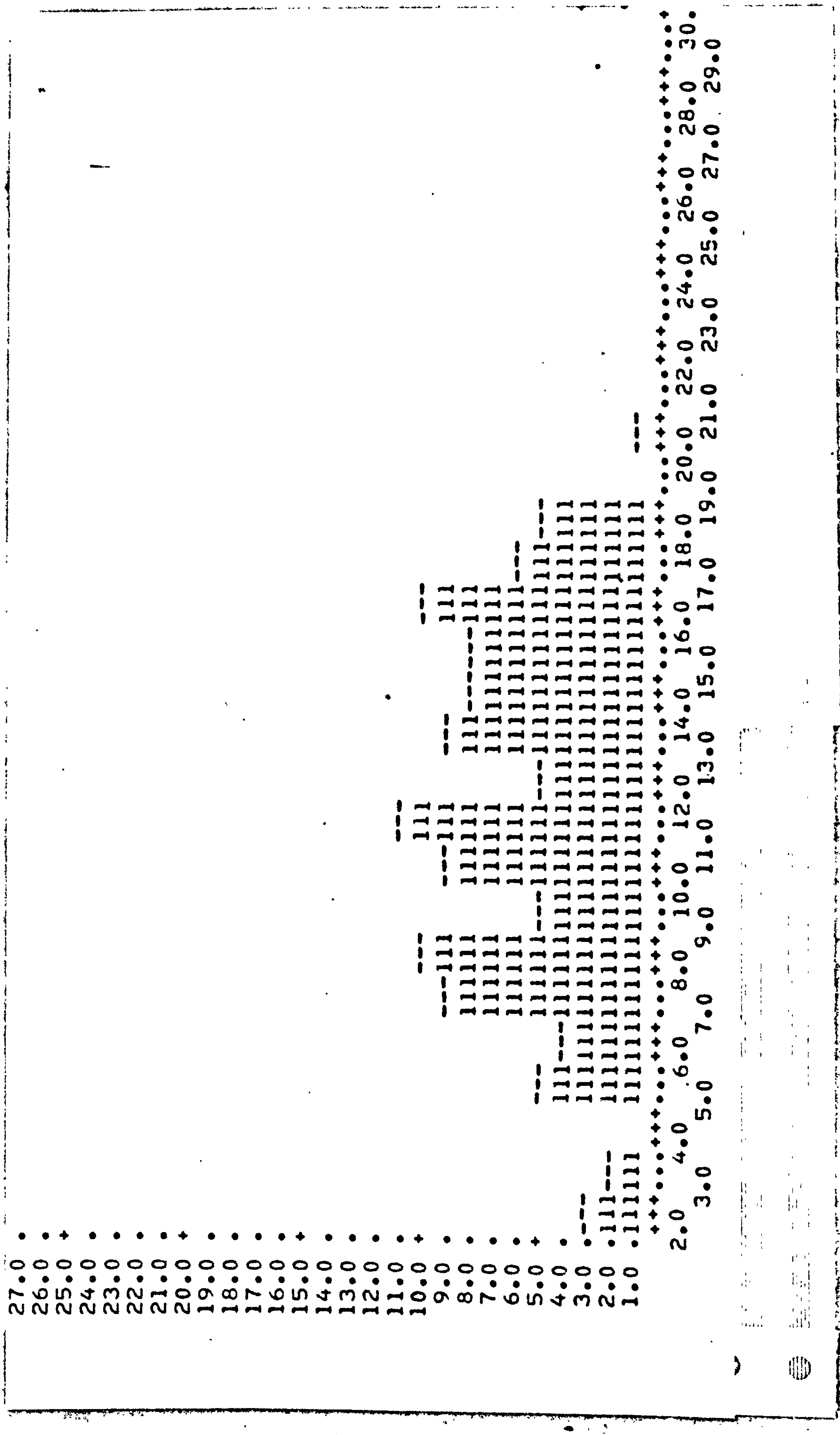


FIG. 15

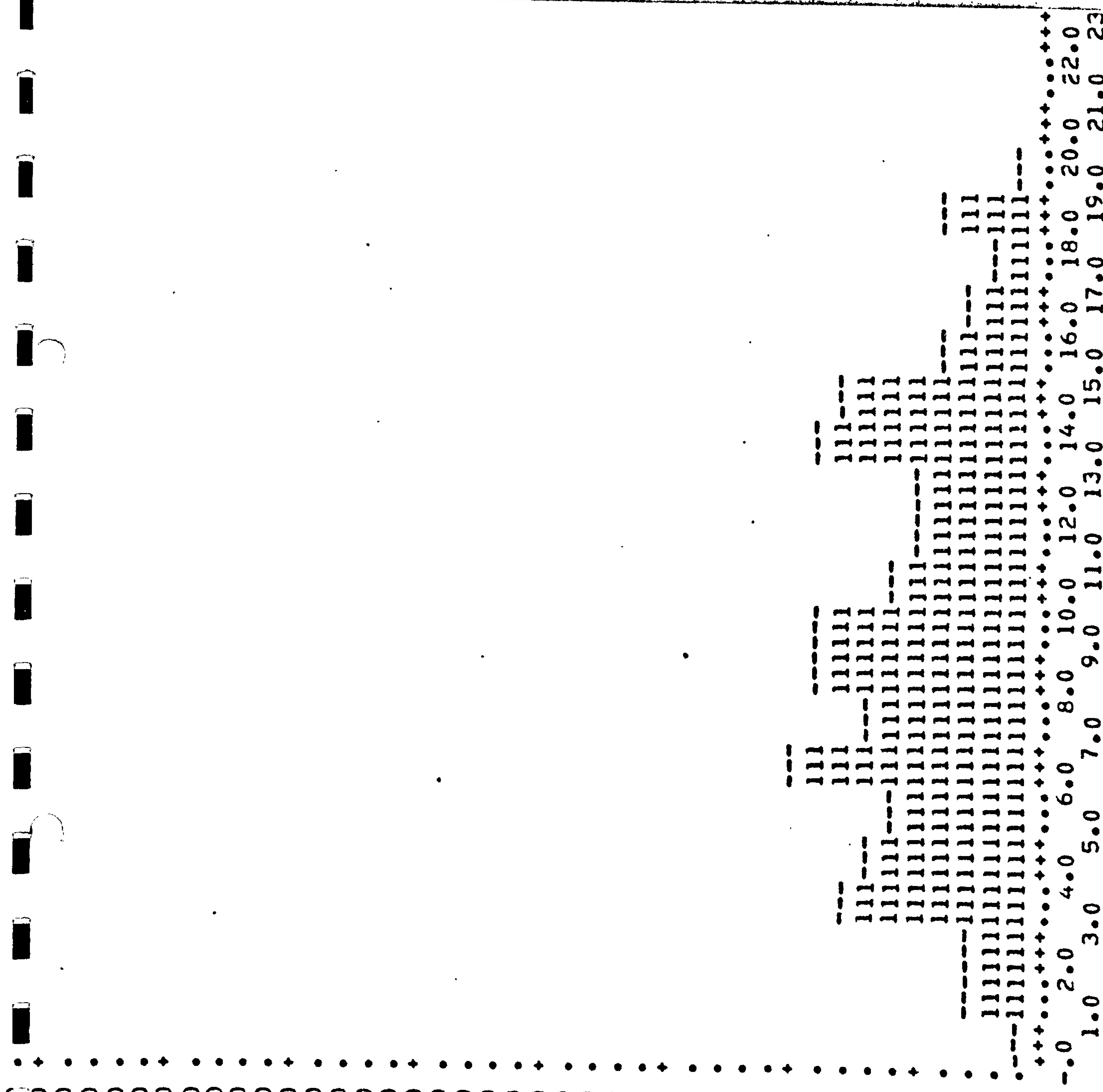

1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0 39.0 40.0
-.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23

FIG. 16

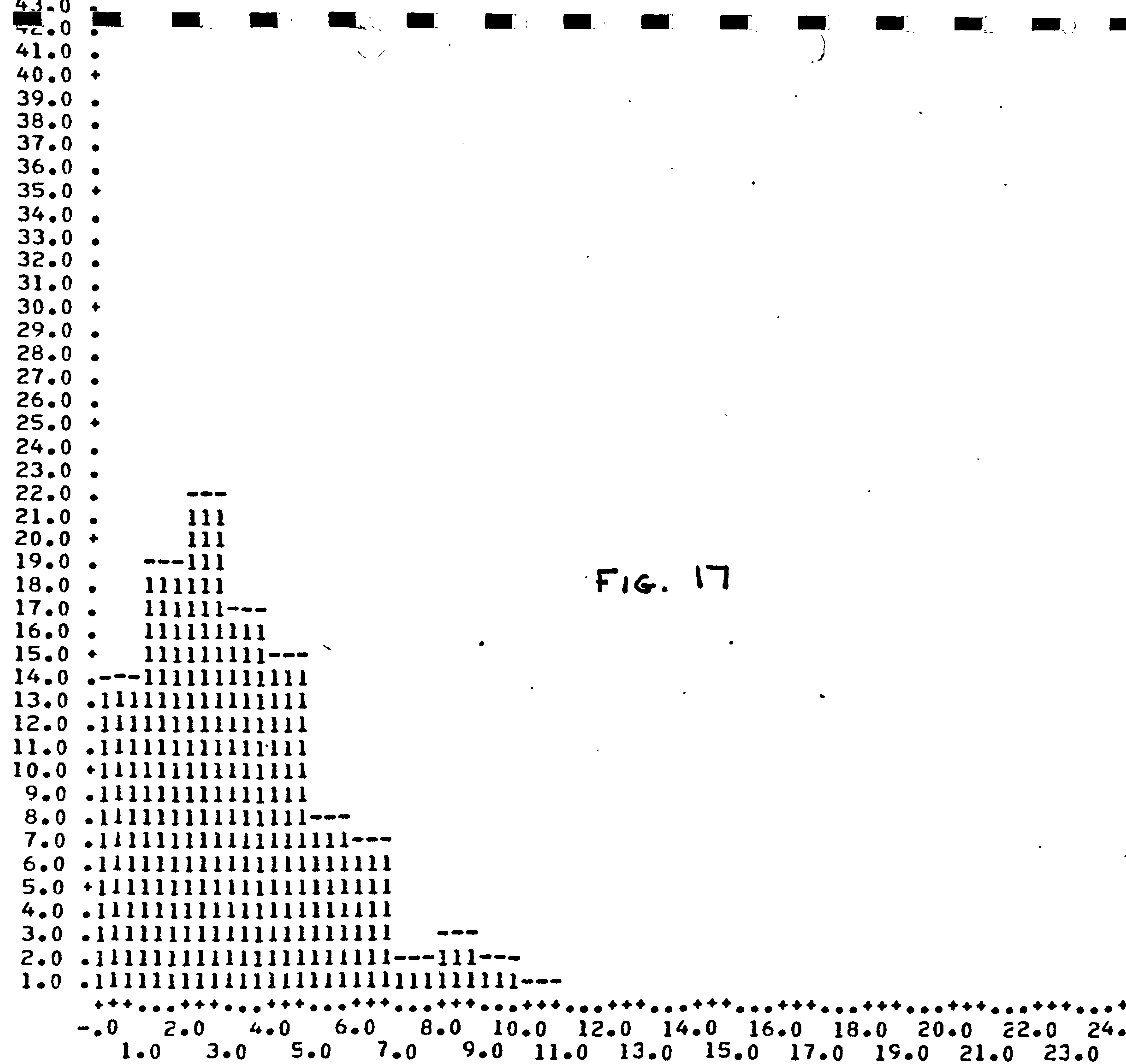


FIG. 17

FIGURES 18 - 30:

These FIGURES show the frequency distributions for the following variables:-

FIGURE 18. Mill Hill Total Score

" 19. Matrices " "

" 20. Mill Hill Total Time

" 21. Matrices " "

" 22. Mill Hill Furneaux Speed Scores

" 23. Matrices " " "

" 24. Mill Hill, Number of Abandonments

" 25. Matrices, " " "

" 26. Choice Reaction Time Slope Coefficients

" 27. 1/1 Reaction Times

" 28. 1/2 " "

" 29. 1/4 " "

" 30. 1/8 " "

18. MH TOT

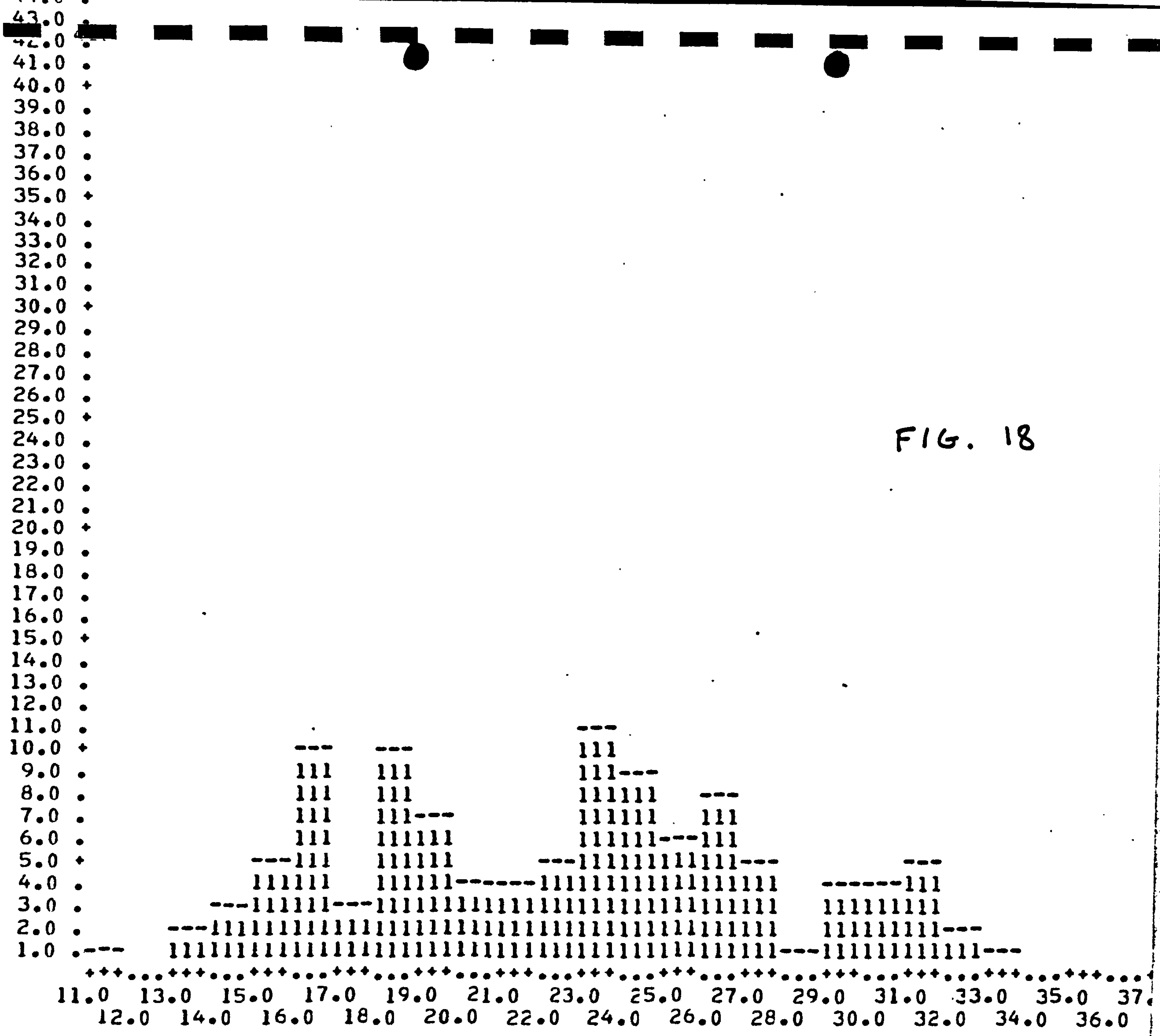


FIG. 18

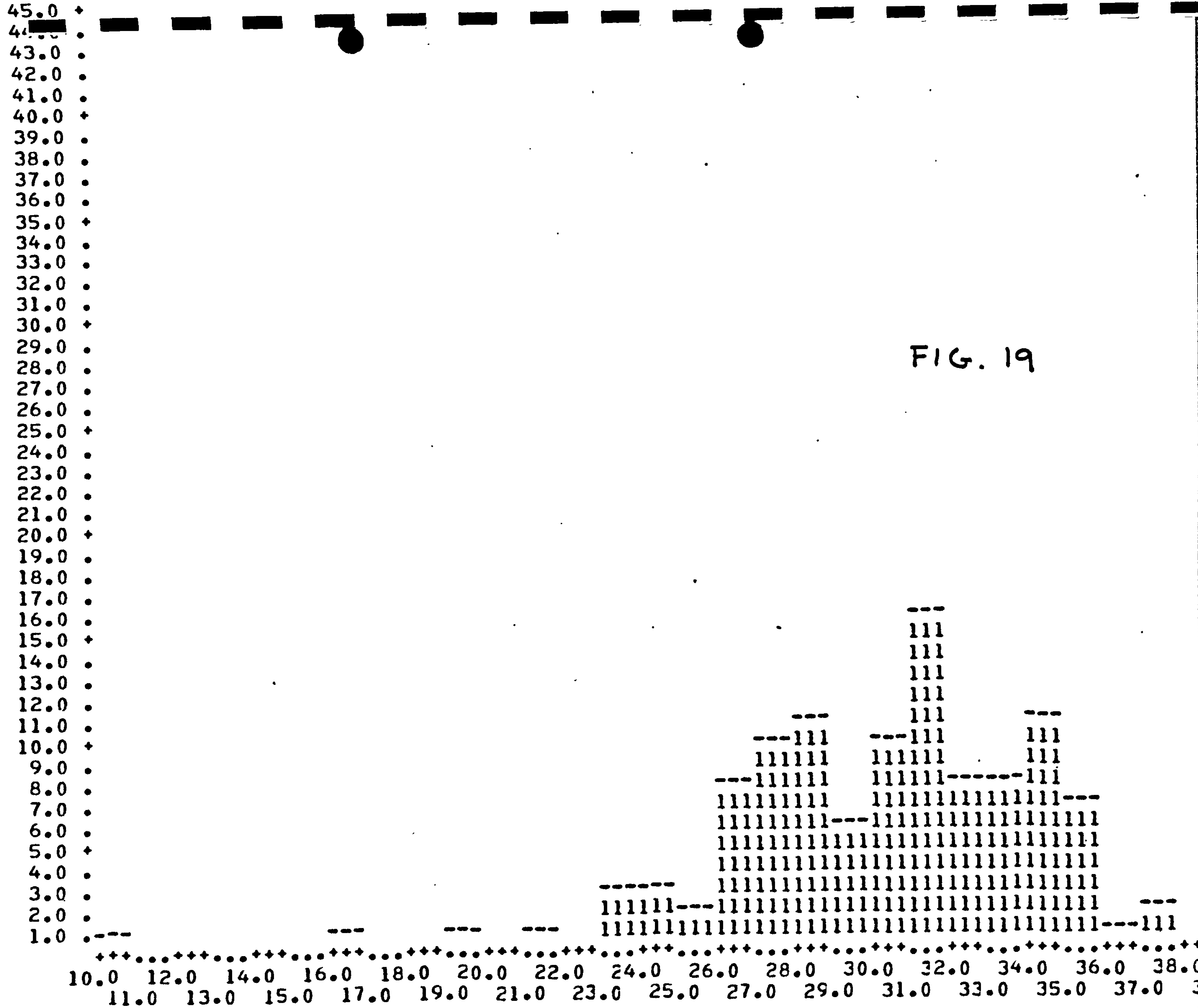

45.0
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
1.0
10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27.0 28.0 29.0 30.0 31.0 32.0 33.0 34.0 35.0 36.0 37.0 38.0

FIG. 19

20 MH TIME.

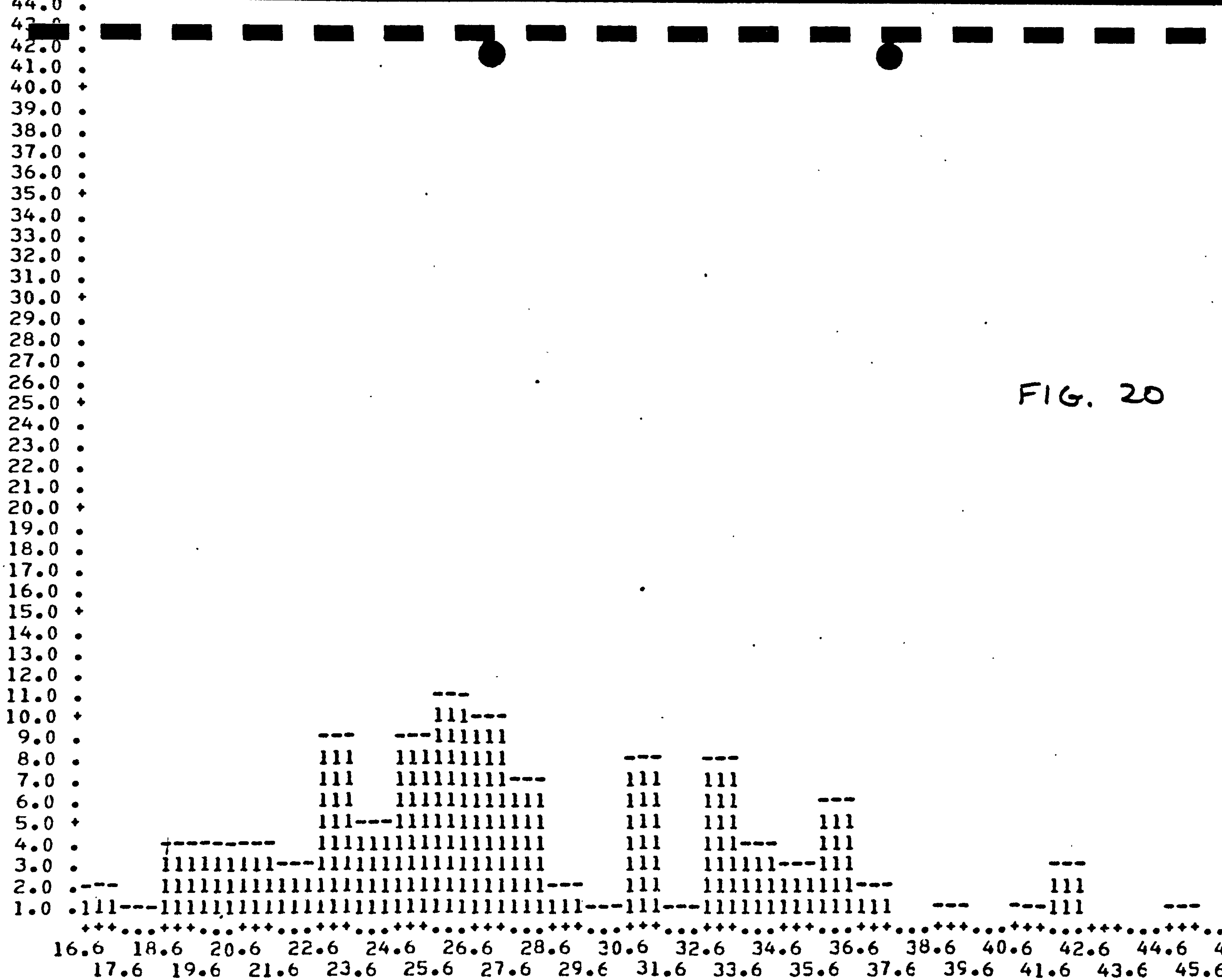


FIG. 20

FIG. 21

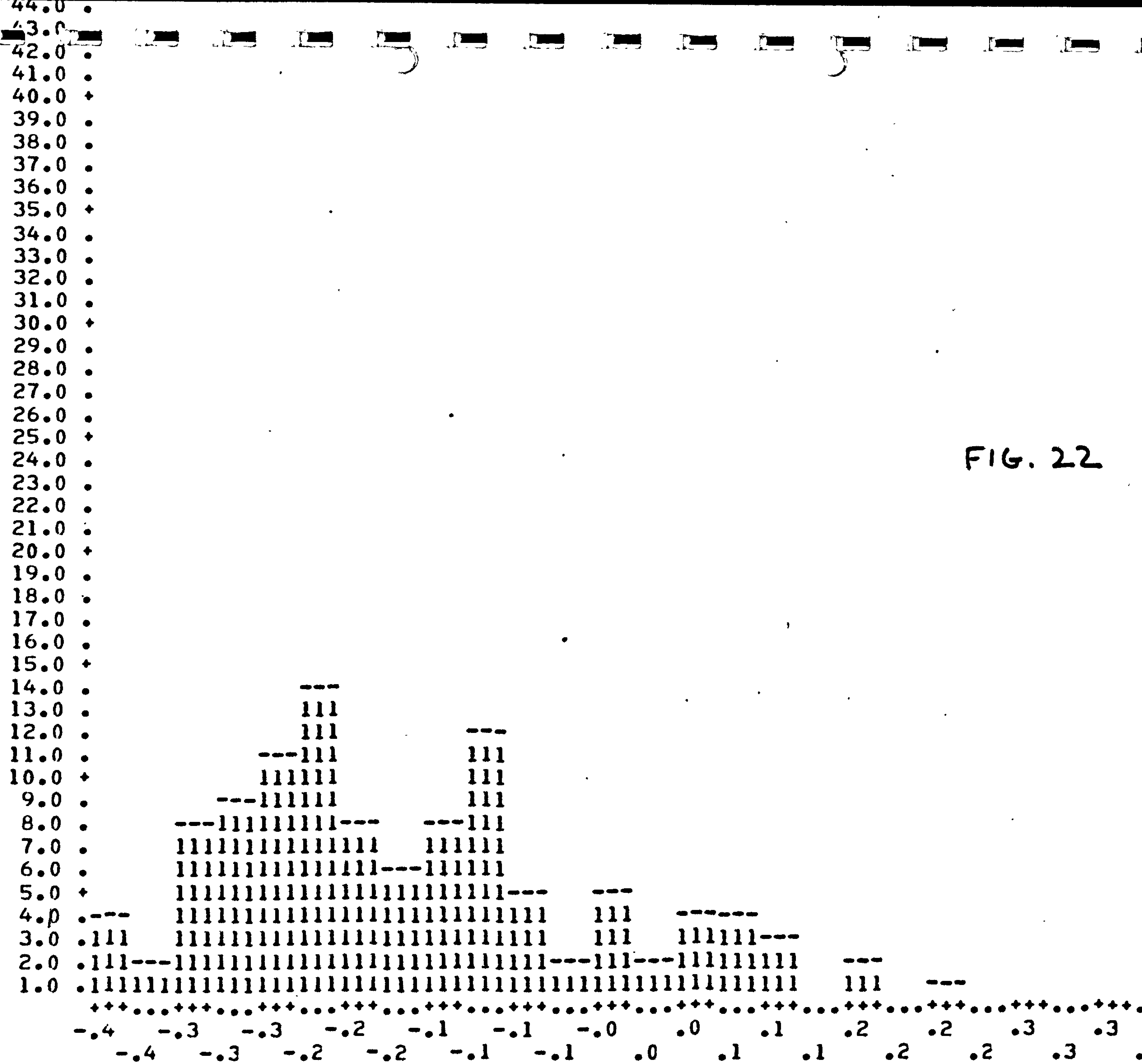
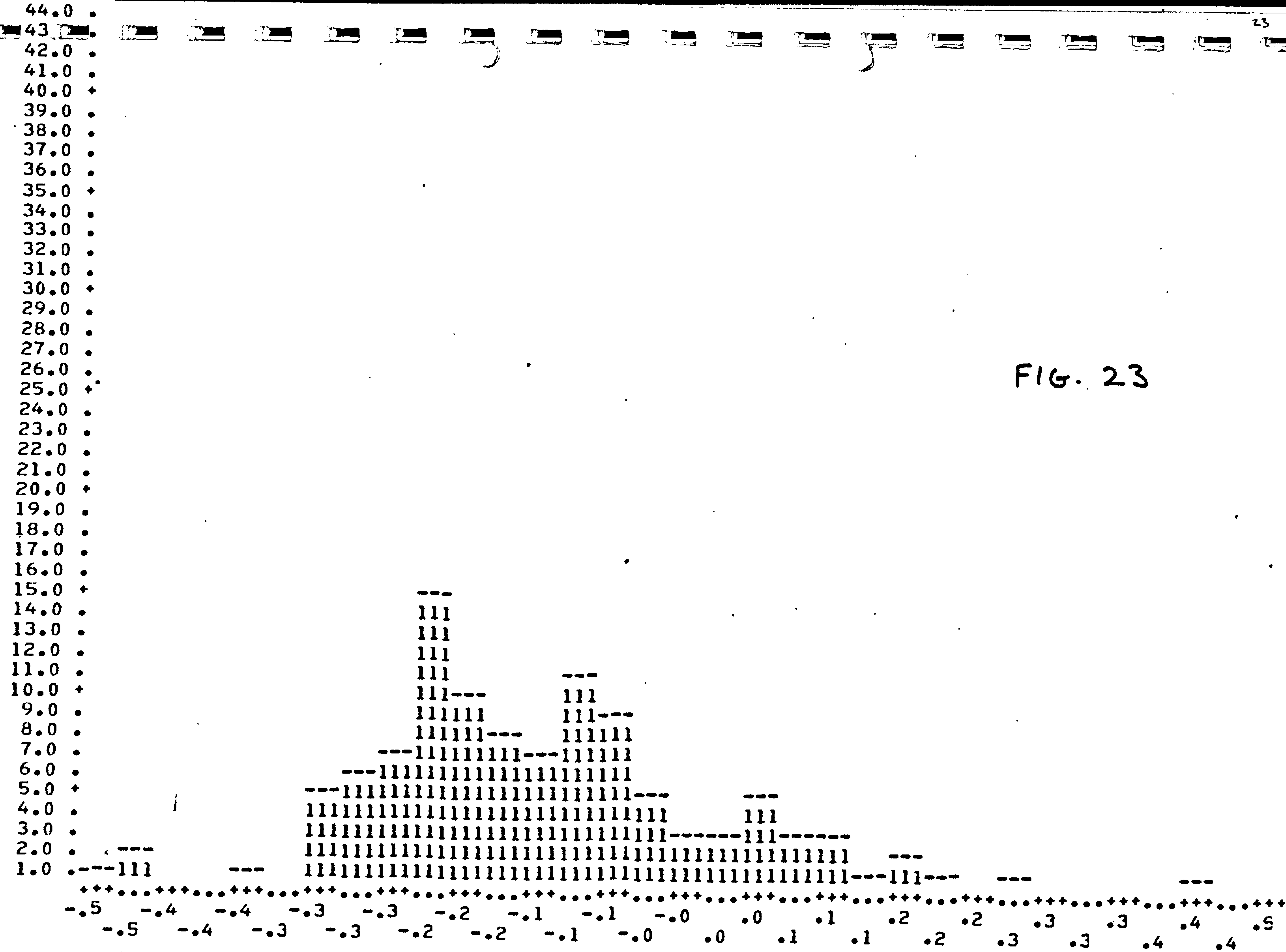

44.0
43.0
42.0
41.0
40.0
39.0
38.0
37.0
36.0
35.0
34.0
33.0
32.0
31.0
30.0
29.0
28.0
27.0
26.0
25.0
24.0
23.0
22.0
21.0
20.0
19.0
18.0
17.0
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
-.4 -.3 -.3 -.2 -.1 -.1 -.0 .0 .1 .2 .2 .3 .3
-.4 -.3 -.2 -.2 -.1 -.1 .0 .1 .1 .2 .2 .3

FIG. 22

FIG. 23


44.0 43.0 42.0 41.0 40.0 39.0 38.0 37.0 36.0 35.0 34.0 33.0 32.0 31.0 30.0 29.0 28.0 27.0 26.0 25.0 24.0 23.0 22.0 21.0 20.0 19.0 18.0 17.0 16.0 15.0 14.0 13.0 12.0 11.0 10.0 9.0 8.0 7.0 6.0 5.0 4.0 3.0 2.0 1.0
-.5 -.5 -.4 -.4 -.4 -.3 -.3 -.3 -.3 -.2 -.2 -.2 -.1 -.1 -.1 -.0 -.0 .0 .0 .1 .1 .1 .2 .2 .2 .3 .3 .3 .3 .4 .4 .4 .5

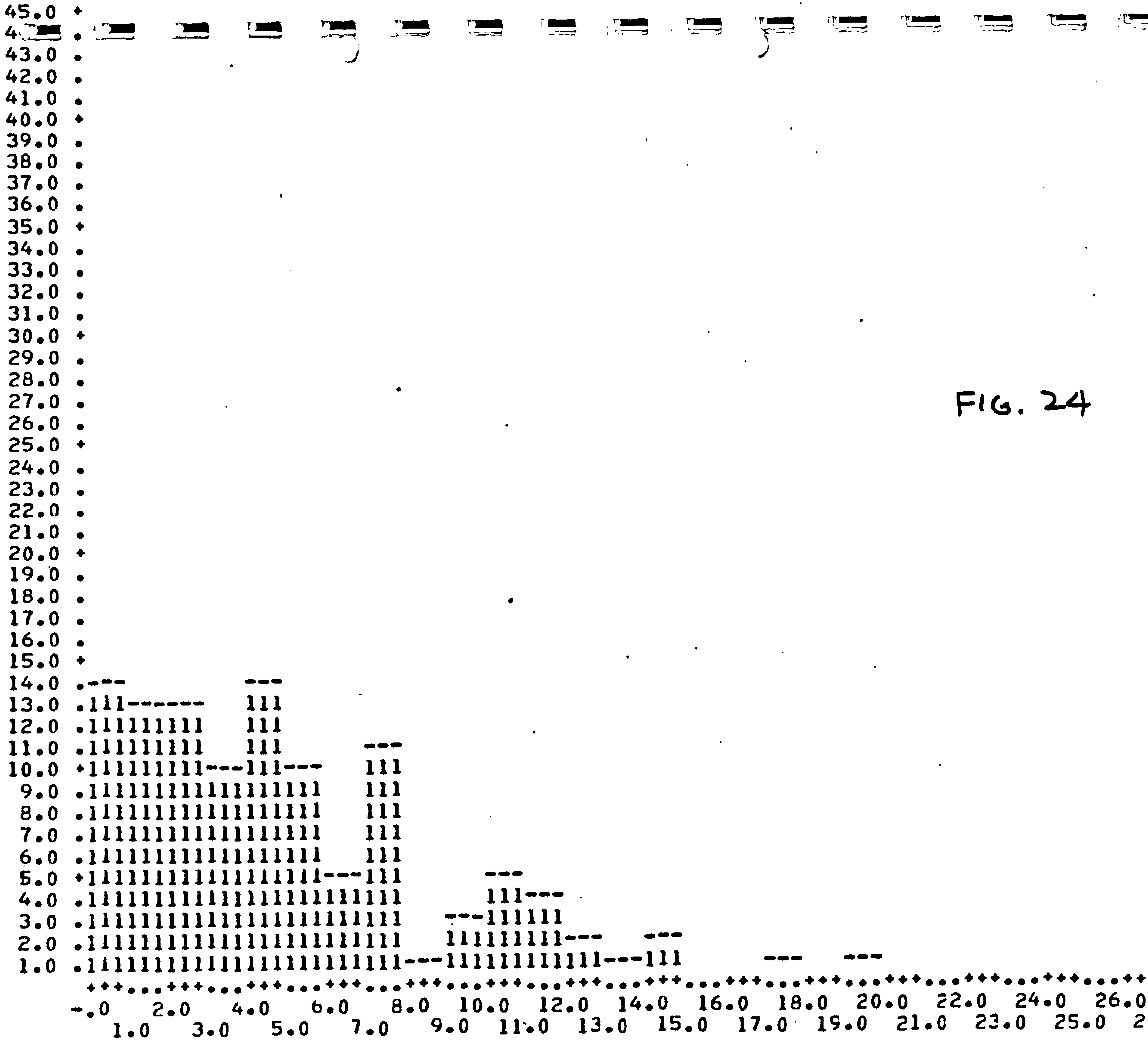

45.0
43.0
42.0
41.0
40.0
39.0
38.0
37.0
36.0
35.0
34.0
33.0
32.0
31.0
30.0
29.0
28.0
27.0
26.0
25.0
24.0
23.0
22.0
21.0
20.0
19.0
18.0
17.0
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
-.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
9.0
10.0
11.0
12.0
13.0
14.0
15.0
16.0
17.0
18.0
19.0
20.0
21.0
22.0
23.0
24.0
25.0
26.0

FIG. 24

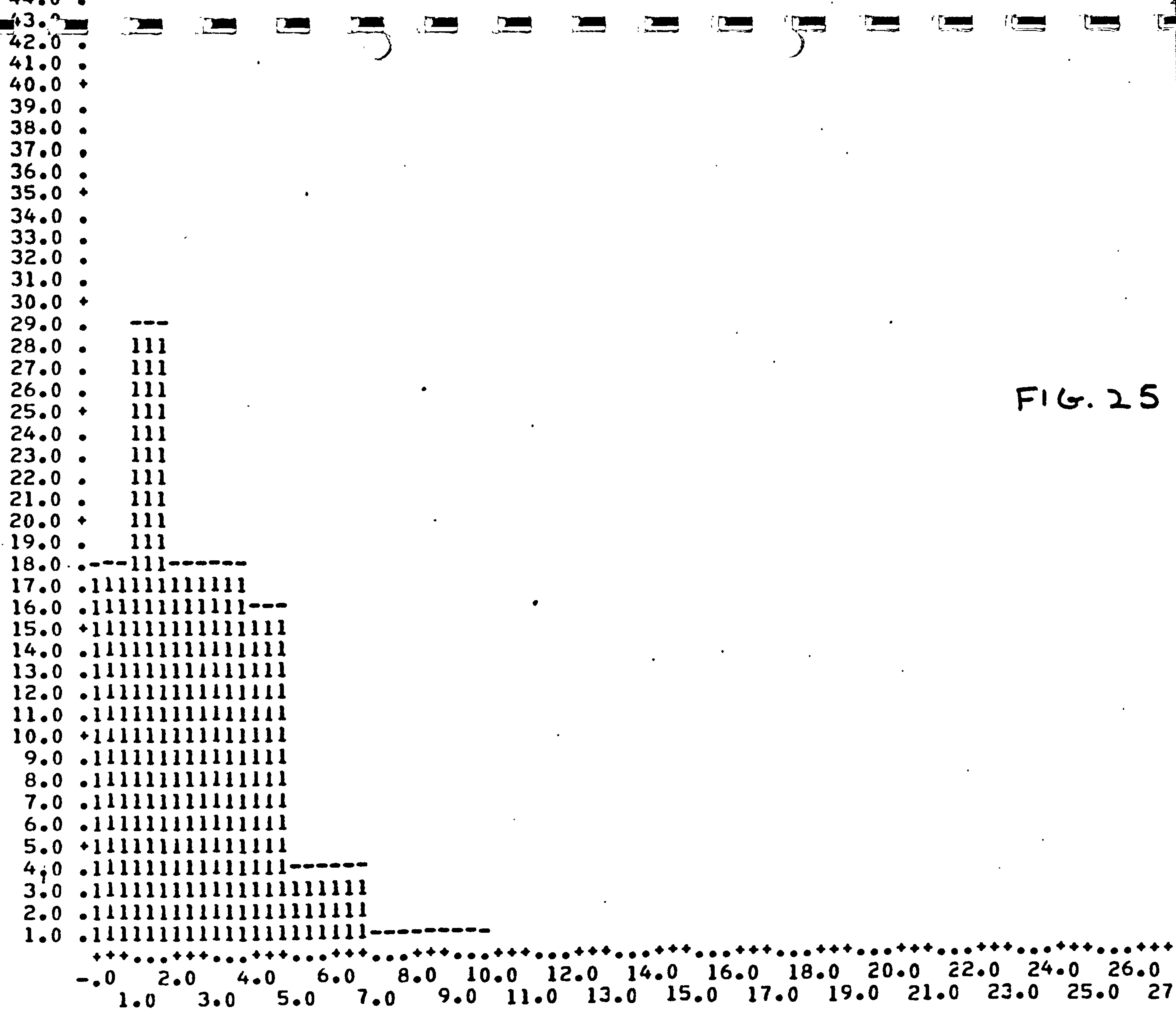
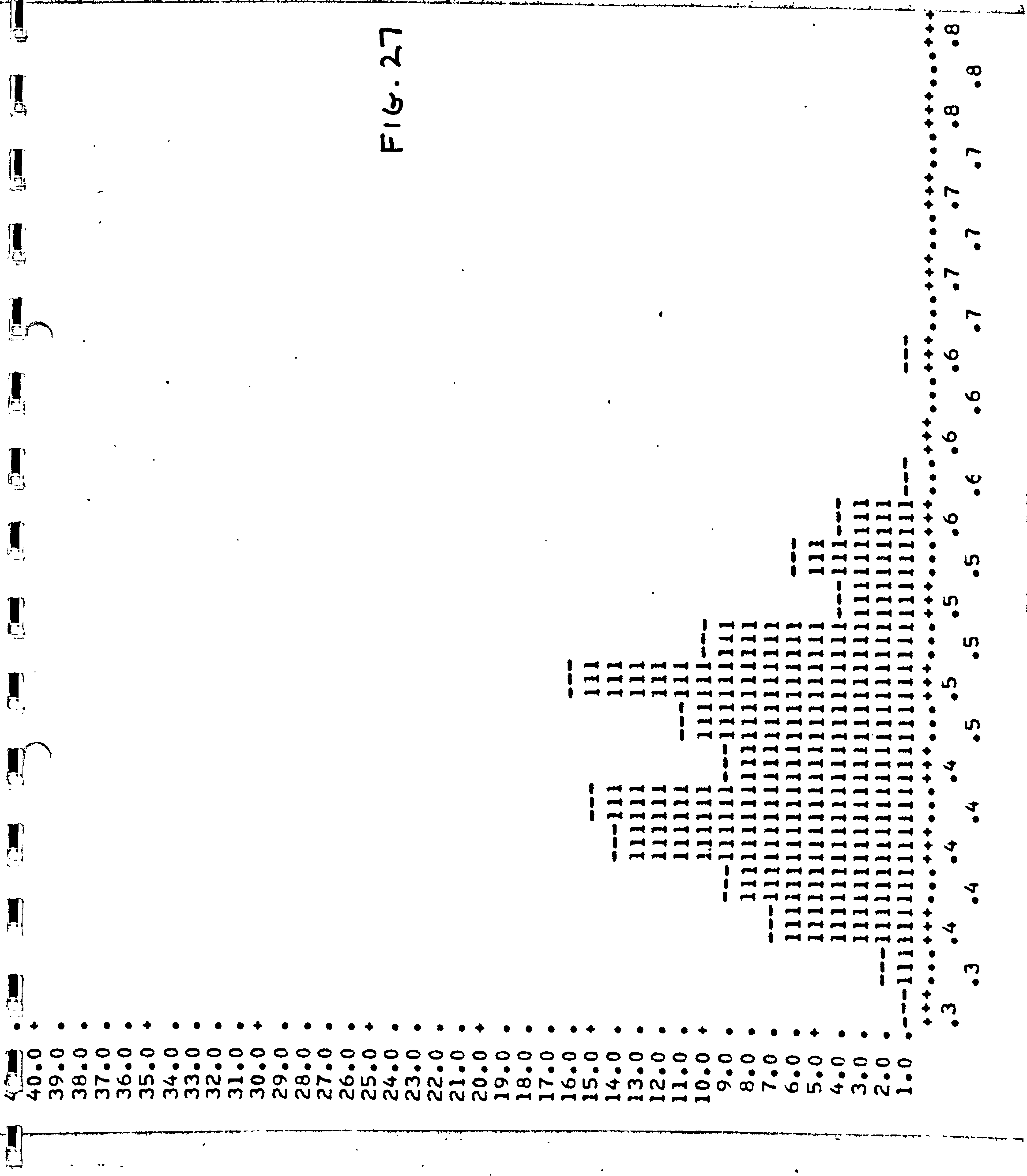

42.0
41.0
40.0
39.0
38.0
37.0
36.0
35.0
34.0
33.0
32.0
31.0
30.0
29.0
28.0
27.0
26.0
25.0
24.0
23.0
22.0
21.0
20.0
19.0
18.0
17.0
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
-.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0 24.0 25.0 26.0 27

FIG. 25

FIG. 26

FIG. 27


40.0
39.0
38.0
37.0
36.0
35.0
34.0
33.0
32.0
31.0
30.0
29.0
28.0
27.0
26.0
25.0
24.0
23.0
22.0
21.0
20.0
19.0
18.0
17.0
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.3 .3 .4 .4 .4 .4 .4 .5 .5 .5 .5 .5 .6 .6 .6 .6 .6 .7 .7 .7 .7 .7 .8 .8 .8

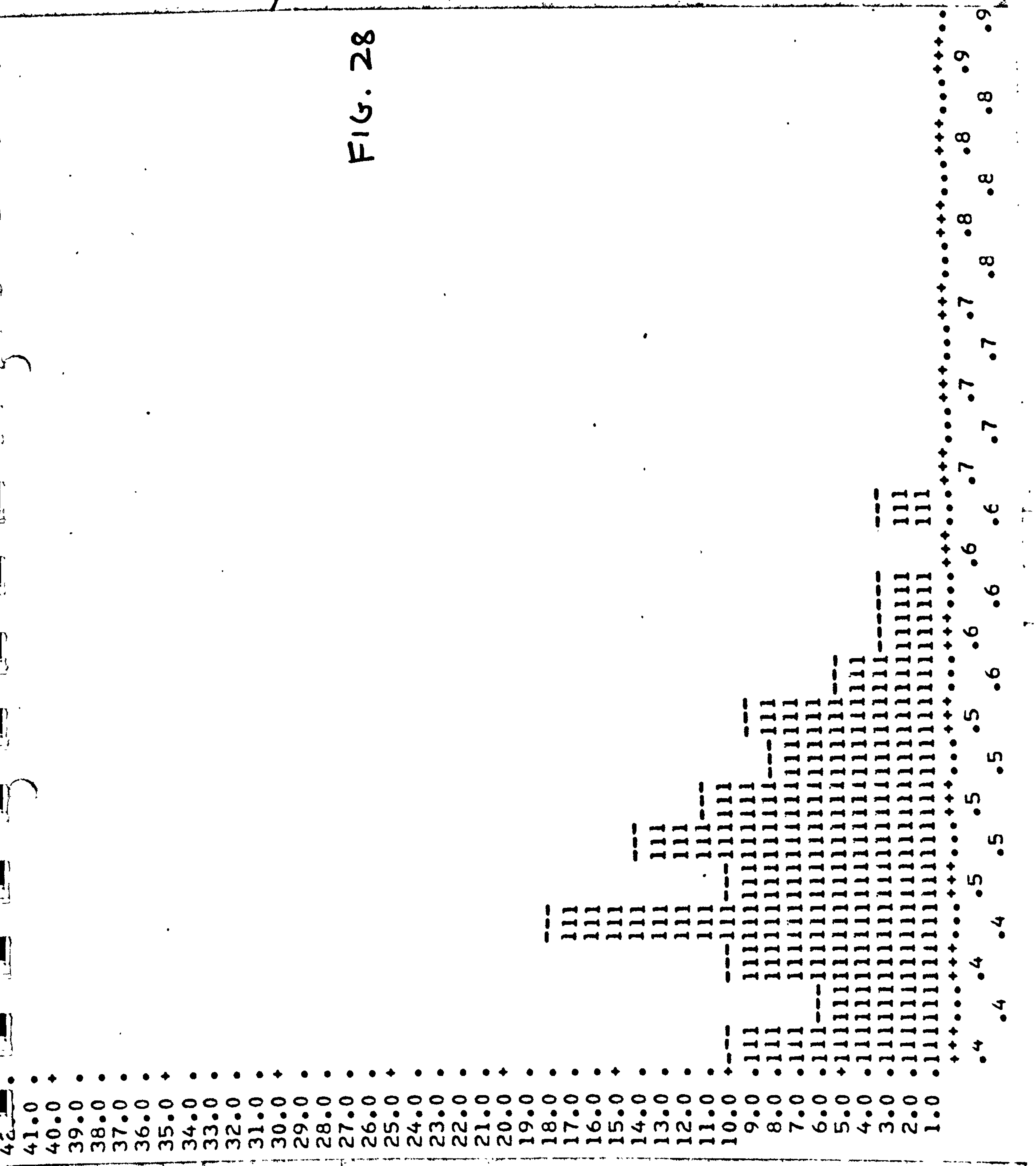
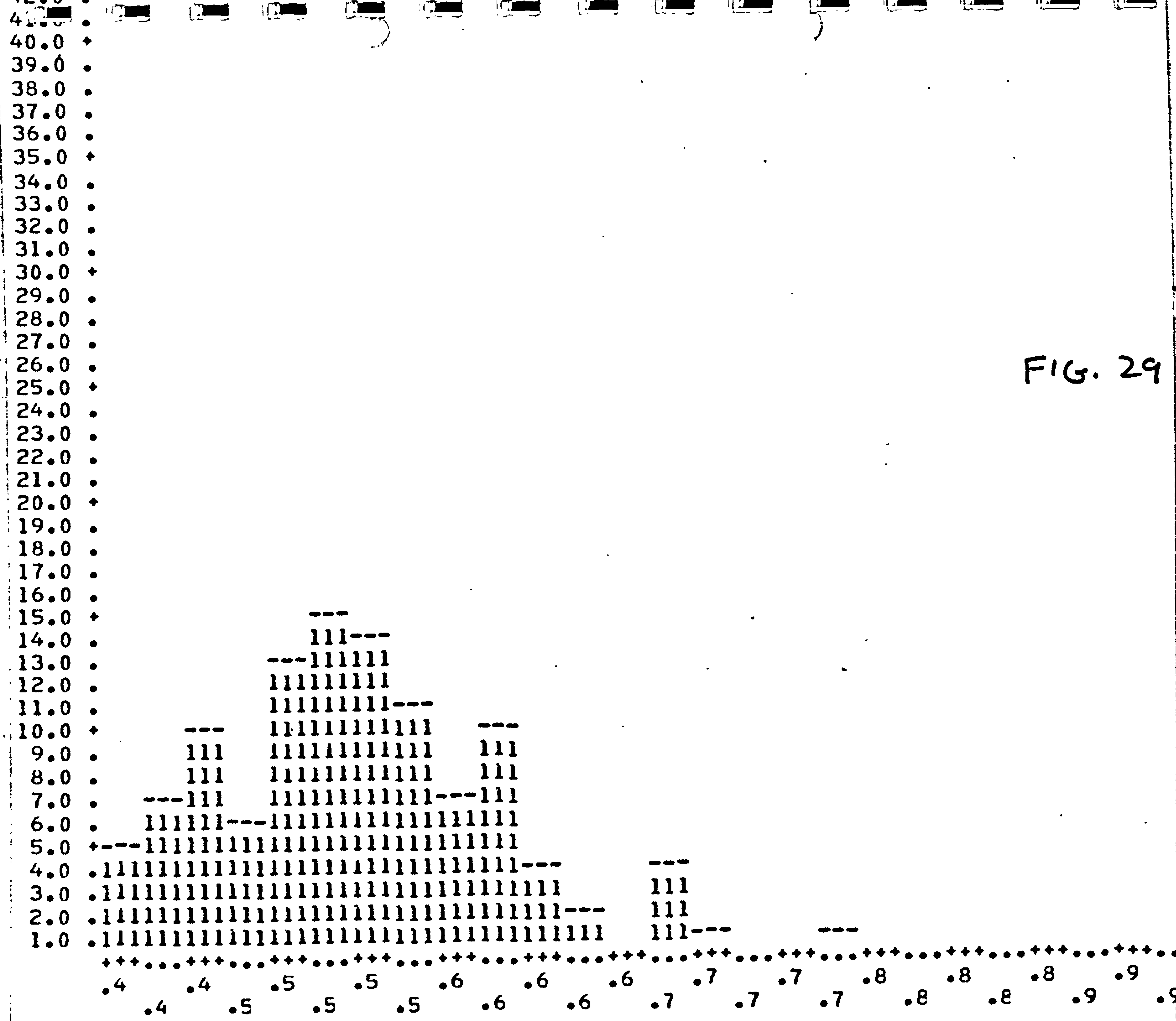


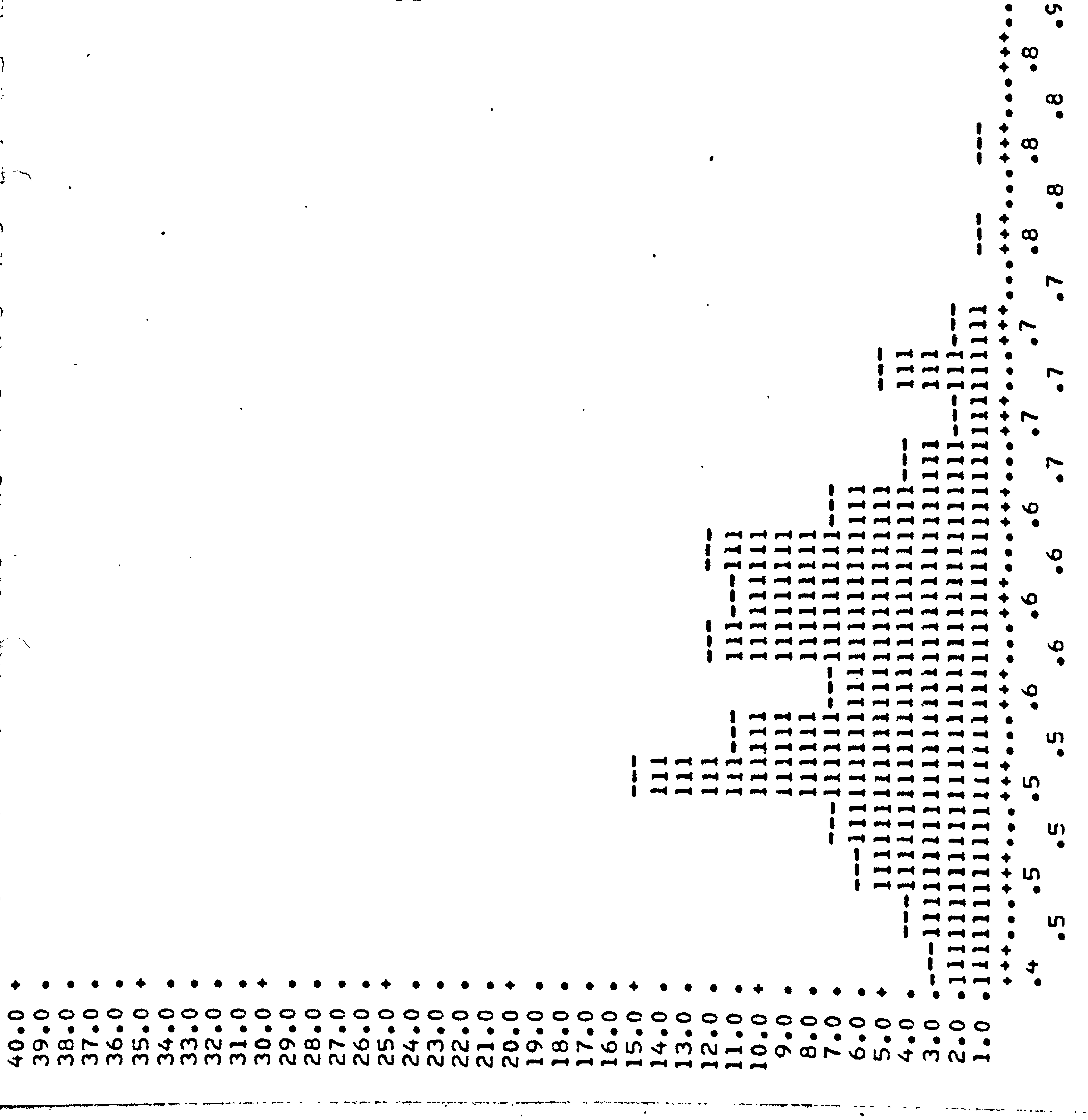
FIG. 28



40.0
39.0
38.0
37.0
36.0
35.0
34.0
33.0
32.0
31.0
30.0
29.0
28.0
27.0
26.0
25.0
24.0
23.0
22.0
21.0
20.0
19.0
18.0
17.0
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.4 .4 .4 .5 .5 .5 .5 .5 .6 .6 .6 .6 .6 .7 .7 .7 .7 .7 .8 .8 .8 .8 .8 .9 .9

FIG. 29

FIG. 30


40.0
39.0
38.0
37.0
36.0
35.0
34.0
33.0
32.0
31.0
30.0
29.0
28.0
27.0
26.0
25.0
24.0
23.0
22.0
21.0
20.0
19.0
18.0
17.0
16.0
15.0
14.0
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
.4 .5 .5 .5 .5 .5 .6 .6 .6 .6 .6 .7 .7 .7 .7 .7 .8 .8 .8 .8 .8 .5 .9 .9 .9

TABLES 11-16:

TABLE 11. Distribution of answers for each of the 6 alternatives on the Mill Hill.
Item Numbers in column 1; Abandonments in column 10.
Frequency of responses to each alternative in columns 2 - 7.

TABLE 12. Mean log. times for each of the answers on the Mill Hill. Row and column identification as for TABLE 11.

TABLE 13. Standard deviations for the log solution times on the Mill Hill. Row and column identification as for TABLE 11.

TABLE 14. Distribution of answers for each of the 8 alternatives on the Matrices.
Item Numbers in column 1; Abandonments in column 10.
Frequency of responses to each alternative in columns 2 - 9.

TABLE 15. Mean log. times for each of the answers on the Matrices. Row and column identification as for TABLE 14.

TABLE 16. Standard deviations for the log solution times on the Matrices. Row and column identification as for TABLE 14

TABLE II.

1	2	3	4	5	6	7	8	9	10
1	0	1	0	0	109	0	0	0	0
2	109	0	1	0	0	0	0	0	0
3	1	106	1	1	0	0	0	0	1
4	0	0	0	110	0	0	0	0	0
5	0	0	7	0	0	102	0	0	1
6	0	0	110	0	0	0	0	0	0
7	107	0	0	2	0	0	0	0	1
8	0	0	0	0	0	110	0	0	0
9	1	108	0	0	0	0	0	0	1
10	0	101	8	0	1	0	0	0	0
11	0	0	109	0	0	0	0	0	1
12	2	0	0	1	104	1	0	0	2
13	0	109	0	0	0	0	0	0	1
14	1	3	10	94	0	0	0	0	2
15	1	7	85	0	0	1	0	0	16
16	103	0	0	3	1	1	0	0	2
17	58	2	11	11	6	0	0	0	22
18	0	0	1	8	14	80	0	0	7
19	1	12	5	69	1	13	0	0	9
20	1	0	7	6	60	9	0	0	27
21	7	58	0	13	1	7	0	0	24
22	0	2	8	4	17	61	0	0	18
23	1	34	30	19	12	5	0	0	9
24	4	32	1	39	7	8	0	0	19
25	1	13	8	2	44	1	0	0	41
26	17	40	7	6	2	0	0	0	38
27	8	3	26	5	7	11	0	0	50
28	9	23	15	34	9	10	0	0	10
29	51	5	3	6	17	4	0	0	24
30	20	8	13	10	17	1	0	0	41
31	1	2	2	6	17	17	0	0	65
32	37	3	11	1	5	1	0	0	52
33	0	21	0	30	2	29	0	0	28

TABLE 12

1	2	3	4	5	6	7	8	9	10
1	0.0000	1.3414	0.0000	0.0000	.7385	0.0000	0.0000	0.0000	0.0000
2	.5754	0.0000	.8899	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	.5250	.6335	1.0546	.3856	0.0000	0.0000	0.0000	0.0000	1.3373
4	0.0000	0.0000	0.0000	.5932	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	.7592	0.0000	0.0000	.7177	0.0000	0.0000	.8519
6	0.0000	0.0000	.6902	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	.6352	0.0000	0.0000	.9571	0.0000	0.0000	0.0000	0.0000	.7356
8	0.0000	0.0000	0.0000	0.0000	0.0000	.6265	0.0000	0.0000	0.0000
9	.6232	.6383	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	.9420
10	0.0000	.6540	.8795	0.0000	.8062	0.0000	0.0000	0.0000	0.0000
11	0.0000	0.0000	.7251	0.0000	0.0000	0.0000	0.0000	0.0000	1.1206
12	.5237	0.0000	0.0000	1.4456	.6369	.9154	0.0000	0.0000	1.0263
13	0.0000	.6482	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0052
14	1.3288	1.0590	.9332	.8157	0.0000	0.0000	0.0000	0.0000	1.5203
15	1.7150	1.0076	.7268	0.0000	0.0000	1.4657	0.0000	0.0000	1.0990
16	.8732	0.0000	0.0000	1.4476	.9186	.9708	0.0000	0.0000	1.1579
17	.8670	1.0685	1.1772	.9747	1.1326	0.0000	0.0000	0.0000	1.0381
18	0.0000	0.0000	1.1926	.8769	.9091	.7839	0.0000	0.0000	1.3873
19	1.2243	1.1885	1.1740	.8633	1.3771	1.0544	0.0000	0.0000	1.2759
20	1.1169	0.0000	1.2014	1.2114	.8190	.9178	0.0000	0.0000	1.2087
21	.9903	.7739	0.0000	1.1962	1.0052	1.2236	0.0000	0.0000	1.1465
22	0.0000	.7801	.8754	1.1521	1.0994	.7784	0.0000	0.0000	1.0253
23	1.4513	.7762	.9459	.9660	1.1677	1.2190	0.0000	0.0000	1.2030
24	.8046	.9579	1.7311	.9520	1.1555	1.1068	0.0000	0.0000	1.1143
25	1.0828	1.0812	1.1293	.7702	.8188	.9253	0.0000	0.0000	.9761
26	1.2322	.8795	1.2585	1.0409	.8111	0.0000	0.0000	0.0000	.9965
27	1.0641	1.3516	1.0162	1.3946	1.2547	1.2315	0.0000	0.0000	1.0611
28	.9361	1.0990	1.0401	.9544	1.0255	1.1511	0.0000	0.0000	1.2385
29	.8685	.9857	1.3811	1.1594	1.2709	1.2467	0.0000	0.0000	1.0705
30	.9803	1.3604	1.1013	1.1026	.8812	.9042	0.0000	0.0000	1.1357
31	1.5027	1.5563	1.0033	1.4176	1.0334	1.0915	0.0000	0.0000	.9636
32	1.0499	1.1121	1.0760	1.2111	1.2898	1.0195	0.0000	0.0000	1.0515
33	0.0000	1.1297	0.0000	1.0959	1.1144	.9518	0.0000	0.0000	1.1433

TABLE 13.

1	2	3	4	5	6	7	8	9	10
1	0.0000	0.0000	0.0000	0.0000	.2010	0.0000	0.0000	0.0000	0.0000
2	.1951	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	.1508	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	0.0000	0.0000	0.0000	.1800	0.0000	0.0000	0.0000	0.0000	0.0000
5	0.0000	0.0000	.3263	0.0000	0.0000	.1479	0.0000	0.0000	0.0000
6	0.0000	0.0000	.1564	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	.2219	0.0000	0.0000	.2339	0.0000	0.0000	0.0000	0.0000	0.0000
8	0.0000	0.0000	0.0000	0.0000	0.0000	.1449	0.0000	0.0000	0.0000
9	0.0000	.2059	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10	0.0000	.2123	.1799	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
11	0.0000	0.0000	.2097	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
12	.2997	0.0000	0.0000	0.0000	.1901	0.0000	0.0000	0.0000	.1081
13	0.0000	.2513	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
14	0.0000	.0562	.3201	.2366	0.0000	0.0000	0.0000	0.0000	.0643
15	0.0000	.2728	.2437	0.0000	0.0000	0.0000	0.0000	0.0000	.2548
16	.2338	0.0000	0.0000	.3302	0.0000	0.0000	0.0000	0.0000	.1322
17	.2552	.0519	.2911	.2897	.3199	0.0000	0.0000	0.0000	.2979
18	0.0000	0.0000	0.0000	.2674	.2420	.2244	0.0000	0.0000	.2692
19	0.0000	.2921	.1675	.2465	0.0000	.2855	0.0000	0.0000	.2274
20	0.0000	0.0000	.2126	.1594	.3043	.1452	0.0000	0.0000	.2785
21	.2396	.2806	0.0000	.3614	0.0000	.2987	0.0000	0.0000	.2101
22	0.0000	.2737	.3474	.4633	.2807	.2145	0.0000	0.0000	.3799
23	0.0000	.2510	.2471	.1888	.3762	.1751	0.0000	0.0000	.2258
24	.2879	.2477	0.0000	.2731	.4243	.2443	0.0000	0.0000	.3179
25	0.0000	.3066	.4248	.1834	.2537	0.0000	0.0000	0.0000	.3083
26	.2160	.3033	.2241	.1939	.0285	0.0000	0.0000	0.0000	.3575
27	.2720	.4140	.2956	.3705	.1775	.3607	0.0000	0.0000	.3021
28	.1270	.3238	.2450	.2977	.2338	.3075	0.0000	0.0000	.1994
29	.2612	.3822	.4057	.3604	.3071	.2075	0.0000	0.0000	.3191
30	.2295	.3714	.3159	.1872	.2325	0.0000	0.0000	0.0000	.1958
31	0.0000	.3819	.1806	.3300	.2209	.3041	0.0000	0.0000	.2697
32	.2673	.1051	.2633	0.0000	.3701	0.0000	0.0000	0.0000	.2696
33	0.0000	.3799	0.0000	.2211	.0900	.2866	0.0000	0.0000	.2401

TABLE 14

1	2	3	4	5	6	7	8	9	10
1	1	0	0	1	0	108	0	0	0
2	0	0	110	0	0	0	0	0	0
3	0	110	0	0	0	0	0	0	0
4	6	102	0	0	0	2	0	0	0
5	110	0	0	0	0	0	0	0	0
6	0	0	105	0	4	1	0	0	0
7	109	0	0	0	1	0	0	0	0
8	17	0	0	2	82	7	0	0	2
9	2	0	0	0	103	4	0	0	1
10	3	6	1	2	94	1	0	1	2
11	71	2	1	9	4	0	11	8	4
12	11	9	1	7	34	9	3	8	28
13	0	0	7	3	3	97	0	0	0
14	104	2	0	1	0	1	0	0	2
15	0	0	25	3	0	79	1	2	0
16	2	4	3	45	12	8	10	11	15
17	2	0	0	105	2	1	0	0	0
18	0	2	100	1	2	0	5	0	0
19	4	1	1	69	11	2	5	1	16
20	7	5	6	2	21	5	5	8	51
21	0	1	109	0	0	0	0	0	0
22	103	1	0	0	1	0	5	0	0
23	3	3	70	0	13	1	3	12	5
24	44	9	11	8	13	1	2	1	21
25	1	2	0	103	2	2	0	0	0
26	0	1	0	1	0	106	0	1	1
27	3	16	4	3	1	4	5	66	8
28	44	3	25	2	0	11	5	2	18
29	2	3	2	4	91	7	0	0	1
30	88	4	1	2	2	0	4	2	7
31	16	1	1	29	0	0	1	49	13
32	10	6	1	7	14	3	1	13	55
33	0	0	0	2	0	108	0	0	0
34	1	0	1	0	0	107	0	0	1
35	110	0	0	0	0	0	0	0	0
36	0	110	0	0	0	0	0	0	0
37	0	2	1	105	2	0	0	0	0
38	0	1	104	2	3	0	0	0	0

TABLE 15

1	2	3	4	5	6	7	8	9	10
1	1.0022	0.0000	0.0000	1.0048	0.0000	.9320	0.0000	0.0000	0.0000
2	0.0000	0.0000	.7659	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	.5769	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	.9350	.8652	0.0000	0.0000	0.0000	.7033	0.0000	0.0000	0.0000
5	.6854	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	.8853	0.0000	.8685	1.2545	0.0000	0.0000	0.0000
7	.8472	0.0000	0.0000	0.0000	.9405	0.0000	0.0000	0.0000	0.0000
8	1.0588	0.0000	0.0000	1.3675	1.2101	1.1639	0.0000	0.0000	1.5929
9	.6954	0.0000	0.0000	0.0000	.9671	.7999	0.0000	0.0000	.9069
10	1.8579	1.4047	.8306	1.4070	1.4935	1.4685	0.0000	1.4840	1.7734
11	1.6357	1.8141	1.4971	1.7932	1.4729	0.0000	1.5463	1.4710	1.7367
12	1.6653	1.9392	1.9253	1.9370	1.9598	1.7223	1.7402	1.7429	1.7364
13	0.0000	0.0000	.9705	1.2178	1.1438	1.0954	0.0000	0.0000	0.0000
14	1.3162	1.1630	0.0000	1.5126	0.0000	1.2106	0.0000	0.0000	1.3974
15	0.0000	0.0000	1.4503	1.5796	0.0000	1.3704	1.0386	1.3461	0.0000
16	1.8099	1.8812	2.0154	1.9518	1.7717	1.8291	1.6123	2.0146	1.9088
17	1.0256	0.0000	0.0000	.9638	.9569	.6532	0.0000	0.0000	0.0000
18	0.0000	1.1321	1.3161	1.2594	1.2920	0.0000	1.4794	0.0000	0.0000
19	1.2721	1.9072	1.7121	1.6250	1.6697	1.5815	1.7586	1.9176	1.7481
20	1.6555	1.9742	1.6834	1.9370	1.9417	2.0002	1.9519	1.7057	1.9549
21	0.0000	.8627	1.0077	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
22	1.2468	1.2114	0.0000	0.0000	1.4415	0.0000	1.3071	0.0000	0.0000
23	1.3895	1.4914	1.5706	0.0000	1.5176	2.0587	1.5669	1.5893	1.4445
24	1.9482	1.8451	1.5834	1.7425	1.8617	1.5963	1.7390	1.3806	1.7630
25	1.1720	1.0425	0.0000	1.0870	1.0999	.8862	0.0000	0.0000	0.0000
26	0.0000	1.4379	0.0000	1.1310	0.0000	1.3737	0.0000	1.2891	1.6608
27	1.3730	1.6489	1.9015	1.8725	.9703	1.6829	1.5874	1.6977	2.0102
28	1.8345	1.9063	1.8296	1.7980	0.0000	1.8295	1.9831	1.1339	1.9850
29	1.5309	1.4505	1.0365	1.3831	1.3827	1.2263	0.0000	0.0000	1.7728
30	1.6659	1.4519	1.9789	1.1508	1.6308	0.0000	1.9474	1.5955	1.7619
31	1.6976	1.5855	1.8107	1.7042	0.0000	0.0000	1.7613	1.7724	1.6204
32	1.7403	2.0359	1.9861	1.8185	1.8963	2.4577	2.2181	1.9457	1.9281
33	0.0000	0.0000	0.0000	.8693	0.0000	.9097	0.0000	0.0000	0.0000
34	1.5347	0.0000	1.5718	0.0000	0.0000	.9315	0.0000	0.0000	.7067
35	.8610	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
36	0.0000	.7109	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
37	0.0000	.4671	1.6190	1.0774	.6238	0.0000	0.0000	0.0000	0.0000
38	0.0000	.9074	.9753	.5320	1.6580	0.0000	0.0000	0.0000	0.0000

TABLE 16

1	2	3	4	5	6	7	8	9	10
1	0.0000	0.0000	0.0000	0.0000	0.0000	.2089	0.0000	0.0000	0.0000
2	0.0000	0.0000	.1898	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.0000	.1820	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	.2835	.2016	0.0000	0.0000	0.0000	.1944	0.0000	0.0000	0.0000
5	.1807	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	0.0000	0.0000	.2000	0.0000	.0525	0.0000	0.0000	0.0000	0.0000
7	.2000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	.3158	0.0000	0.0000	.2610	.2379	.4192	0.0000	0.0000	.1533
9	.1984	0.0000	0.0000	0.0000	.2388	.2687	0.0000	0.0000	0.0000
10	.4992	.2850	0.0000	.2180	.2245	0.0000	0.0000	0.0000	.0476
11	.2143	.1744	0.0000	.3364	.2749	0.0000	.2465	.2245	.2457
12	.3841	.2993	0.0000	.3188	.2119	.2741	.3753	.2923	.3553
13	0.0000	0.0000	.0814	.2924	.1008	.2145	0.0000	0.0000	0.0000
14	.2179	.3683	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	.0656
15	0.0000	0.0000	.2868	.2162	0.0000	.2471	0.0000	.3227	0.0000
16	.6813	.1686	.1546	.2483	.3316	.2904	.3115	.3117	.3003
17	.0252	0.0000	0.0000	.2235	.3313	0.0000	0.0000	0.0000	0.0000
18	0.0000	.1452	.2543	0.0000	.1950	0.0000	.3296	0.0000	0.0000
19	.2855	0.0000	0.0000	.2620	.2577	.0733	.2387	0.0000	.2167
20	.2229	.4743	.1713	.3719	.2917	.4383	.4736	.5576	.2491
21	0.0000	0.0000	.2301	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
22	.2497	0.0000	0.0000	0.0000	0.0000	0.0000	.2383	0.0000	0.0000
23	.1463	.1070	.3125	0.0000	.3192	0.0000	.1831	.2999	.3894
24	.2223	.1809	.3930	.3140	.2824	0.0000	.3920	0.0000	.3083
25	0.0000	.1600	0.0000	.2362	.0171	.0307	0.0000	0.0000	0.0000
26	0.0000	0.0000	0.0000	0.0000	0.0000	.1960	0.0000	0.0000	0.0000
27	.1529	.3027	.3897	.1238	0.0000	.2743	.2751	.2906	.2441
28	.2557	.1803	.2258	.1374	0.0000	.2321	.1941	.3451	.2323
29	.6618	.1671	.1594	.1990	.2942	.3834	0.0000	0.0000	0.0000
30	.2277	.2516	0.0000	.1695	.3375	0.0000	.2605	.4636	.1554
31	.2667	0.0000	0.0000	.1932	0.0000	0.0000	0.0000	.1949	.4329
32	.2683	.3187	0.0000	.3998	.3581	.2811	0.0000	.3124	.2801
33	0.0000	0.0000	0.0000	.2609	0.0000	.1840	0.0000	0.0000	0.0000
34	0.0000	0.0000	0.0000	0.0000	0.0000	.2544	0.0000	0.0000	0.0000
35	.2195	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
36	0.0000	.2093	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
37	0.0000	.1053	0.0000	.1822	.4443	0.0000	0.0000	0.0000	0.0000
38	0.0000	0.0000	.2471	.1006	.3246	0.0000	0.0000	0.0000	0.0000

63	P.E.N. incomplete
68	Apparatus fault
112	Reaction times lost due to failure to respond.

It is interesting to note that the "apparatus fault" was due to the experimenter's failure to reset the slide projector to the starting position. This problem is readily overcome by a small sub-routine which would step the magazine to the appropriate slide in preparation for the next subject.

The incomplete P.E.N's were due to a failure to check that both sides had been filled in. One P.E.N. was not printed correctly and this error and the others were not noticed by the receptionist whose responsibility it was to gather these data from subjects.

The basic data on the 110 subjects are presented in Table 10. Distributions of the subjects on the main variables are presented in Figs. 14 to 17.

2. Preliminary Results:

The results given in this section are the preliminary results of this study. A detailed presentation of the data will be given in a later chapter.

Tables 11 to 16 present the distribution of responses for each test and the means and standard deviations for each solution in the group of 110 subjects.

Figs. 18 to 30 show the distributions for the basic study variables.

3. Item Difficulty : The Mill Hill:

The procedures for scaling items in terms of their time difficulty are detailed by Furneaux in his 1961 paper. Despite the detail his exposition does not permit a direct application for scaling new test material. The lack of clarity will be apparent as this chapter proceeds. Brierley (1969) also noted some of the obscurities in Furneaux's statements and adopted his own interpretation when scaling items. The present author has consulted both these sources in attempting to scale the Mill Hill and Advanced Matrices data.

In this chapter, the details of the analysis presented by Furneaux will be summarised for each major step. This will be followed by a statement of outcome in the present study. This detail is necessitated by the lack of clarity in Furneaux's description. It will enable the reader to evaluate the validity of the present procedures. Such detail should also facilitate any later attempt to utilize Furneaux's procedures. Finally, as will be seen later, this approach makes apparent some of the problems in Furneaux's procedures.

There will be no attempt to reproduce the raw data for this part of the study as they would occupy too much space. Altogether, on the 110 subjects there were 15,620 units of data (71 test items producing 142 units per subject - i.e. solution alternative, solution time). All of Furneaux's procedures were set up as computer programmes written by the author. It is not possible to accomplish a full Furneaux scaling in one pass through the computer. A number of separate programmes must be used as the output has to be examined at each stage. Every programme was thoroughly checked before the output was used.

In the course of his presentation, Furneaux noted that raw solution times did not satisfy the assumptions necessary for the further development of his approach. He therefore used a log 10 transformation. It should be noted that while he adopted a different symbol for the log time measures (T_r for logs, t_r for raw times) he did not use the log notation consistently. In the presentation which follows, all times are in their log 10 form.

<u>Furneaux : Stages I and II</u>:

1. Compute the proportion of correct solutions for all items for all subjects.
2. Select a group of items with the lowest conventional difficulty. These are called the REFERENCE items.
3. For all subjects who have solved at least half of the reference problems, compute mean correct solution time on the reference problems.
4. For subjects who have abandoned <u>any</u> item, compute mean abandonment time for <u>all</u> abandoned items. (Procedurally, this is the most useful first step and is the only one which does not have to be repeated).
5. Compute mean abandonment time for the group and divide subjects into high and low continuant subjects (i.e. above and below mean, respectively).
6. Divide <u>high</u> continuant subjects into three groups in terms of their mean time to correct solution for the reference problems.
7. "Group F" subjects are those who show the fastest mean time to correct solution on the reference problems; "Group M" is the designation of the moderately fast subjects.
8. Unless otherwise specified, all subsequent computations are carried out on data from Group M. With ready accessability to a computer, it is useful to carry out the computations for all subjects, especially those in Group F as they are required at a later stage. This is a simple programming task.

9. Compute the mean correct solution time for each reference problem for Group M. These times are taken as a first approximation of the time difficulties of the items (D_{Tr}).

10. Compute a speed constant for each subject. From each correct reference problem, subtract the D_{Tr} for that problem. Add these differences and divide by the number of differences. The result is the K_{Tr} (speed constant) for each subject. (K_{Tr} is thus an average deviation score for correct solutions. The more negative the value, the faster the subject on the reference items).

11. Compute a critical difficulty value for each subject (c D_{Tr}). Select the smallest abandonment time on any item for the subject. From this value subtract the largest difference found in 10 above. The resulting value is a first approximation to the critical difficulty for that subject.

12. Two checks should now be carried out.

    A. It is necessary to ensure that the lowest value of c D_{Tr} for any subject is greater than the highest D_{Tr} for any reference item. Furneaux does not specify what action should be taken if this test is failed. It does however indicate that the reference items were ambiguous inputs and that a new reference set should be chosen and steps 1 to 12A repeated.

    On the basis of this writer's experience, it is essential that this part of the scaling be satisfactorily completed before attempting 12B. Failure to do so results in unnecessay further computations.

    The aim at this stage should be to maximize the number of items and subjects for whom these items are unambiguous inputs. A useful initial guide in selecting the reference items is their conventional difficulty, but this did not produce fully satisfactory results in the present study. Additional useful information is provided by mean and variance of correct time : items of low mean time and variance tend to lead to a more rapid solution of Stages I and II scaling.

    B. The relationship between item difficulties in Groups M and F should be linear. Again Furneaux does not specify what corrective action should be taken.

The Present Study : Stages I and II:

The computations specified by Furneaux were carried out on 10 conventionally 'easy' items. Although the Group M and F item difficulties showed the predicted linear relationship, few subjects satisfied the test in 12A above, suggesting that these items were ambiguous inputs and therefore

TABLE 17. Time difficulties for the reference items on the Mill Hill for groups M and F. Based on 12 and 25 subjects, respectively.

ITEM	$D_{Tr}(M)$	$D_{Tr}(F)$
1	.8697	.6600
2	.7643	.5190
3	.7464	.5985
4	.7634	.5241
5	.8409	.6697
6	.8343	.6284
8	.8032	.5671

TABLE 18. Critical difficulties for groups M and F at the end of Stage II, Mill Hill.

GROUP M

SUBJECT	$_{C}D_{Tr}$
44	.9151
50	.9266
87	.9468
107	1.2432
85	.9646
104	1.0838
54	.9571
76	1.3034
109	1.0448
29	1.1993
77	1.0550
47	1.0860

GROUP F

SUBJECT	$_{C}D_{Tr}$
113	1.2046
67	1.0411
62	1.1264
73	.8105
19	.8361
117	.9847
53	1.1241
34	.9440
114	.9485
92	1.0822
36	1.0580
82	.9156
27	.7980
101	.8436
81	.8115
35	.7518
66	.7329
86	.8170
14	.6937
93	.9239
28	.6914
91	1.0015
23	.7065
115	.8011
18	.8050

inappropriate as reference items. Furneaux foresaw this as a possible problem but appears from his report not to have encountered it. However, his foresight might well have been a consequence of his encountering similar problems. Brierley (1969) does not mention any problems at this point, but he had adopted a different initial procedure, dividing his small group of subjects into an M and an F group. He did not first select continuant subjects, presumably becuase of the small sample.

Following this initial failure, a number of attempts to obtain a satisfactory set of reference items was made. Computationally, an order of proceeding different to Furneaux's was adopted, beginning with step 5 which identifies the high continuance subjects.

The mean abandonment time was 1.0739 log secs. However, not all subjects with abandonment times above the cut-off can be included in subsequent analyses : the later procedures require the variance in abandonment times to be computed. Therefore, high continuant subjects who abandoned 1 or less items have to be excluded at this stage. 50 subjects remained following this initial analysis.

The distribution of abandonments is presented in Fig. 24. From this it can be seen that data from 27 subjects (0 or 1 abandonments) could not be used in the initial scaling. This was the first sign of a major problem in Furneaux's approach, the high rate of data loss as the analysis proceeds. This problem will be considered in detail in a subsequent chapter.

The sub-division of continuant subjects used in this study did not follow Furneaux's procedures exactly. With only 50 subjects available, it was decided to divide them into F and M groups of 25 subjects each. It will be recalled that Furneaux divided his subjects into 3 groups (the F and M groups containing 40 subjects). It was felt that if Furneaux's procedure was adopted, too few data would be available for later analyses. The present procedure also differed from that used by Brierley (1969). He omitted the step which requires the subjects to be divided into high and low continuant groups, thereby departing as well from Furneaux's specifications.

The achievement of Furneaux's criteria for Stages I and II (steps 12A and B) was accomplished only after extended trial-and-error. Several attempts were made before it was possible to satisfy these criteria for 7 items (No's. 1, 2, 3, 4, 5, 6, and 8), and even then, the data for only 12 group M subjects could be used. The time difficulties for these items are given in Table 17 for Groups M and F. Critical difficulties for the subjects in these groups are presented in Table 18.

From these Tables, it can be seen that the greatest difficulty for a Group M reference item was .8697 and the lowest critical difficulty for any subject was .9151. For Group F the respective values were .6697 and .6914. These data satisfy step 12A of Stages I and II.

The second test is more problematic. With only 7 items and 12 and 25 subjects, it is inappropriate to use regression analysis to determine the relationship between the two sets of item difficulties. Furneaux did not record how many items were involved in his analysis. Brierley (1969) used only 10 items. However, as this was only a preliminary test on a first approximation, the present writer accepted the precedent of the previous workers.

The regression coefficient was computed using the B.M.D 05R programme (Dixon 1970) for polynomial regression analysis for a 4th degree polynomial. The data from this analysis are presented in Table 19. For this analysis, data from Group M were used as predictors.

The regression analysis produced a linear coefficient of 1.05 (s.e. 0.35, r = .80) which compares favourably with Furneaux's finding (linear coefficient 1.08) and with that reported by Brierley (.96). Neither of these writers tested for higher orders of curvilinearity. The present results indicate major linear and cubic components and visual inspection of Brierley's data also suggests that a cubic component was present.

While the statistical limitations of these analyses must be emphasized, the results appear adequate at this point, so that the next stage of scaling can proceed. The only reservation concerns the significant cubic component, but its implications are not very clear. As the difficulty values are only approximations which will be refined in later stages of the analysis, a significant cubic component does not at this point invalidate Furneaux's main conclusion : it would only do so if a substantial cubic component was found in the relationship between item difficulty and speed at the completion of scaling.

Furneaux : Stage III:

1. For every subject, compute speed constants using correct response times for reference items. (Repeat of steps 9 and 10, Stages I and II). What is not explicit at this point is how the necessary critical difficulties that are needed should be computed. Either one can use the D_{Tr} values already computed for Group M on the reference problems, or, they can be recomputed from the data that are available for all subjects. The solution adopted in the present study is described in the next section.

TABLE 19. Results of a 4th degree orthogonal polynomial analysis on item difficulties for Groups M and F. Item difficulties x 100.

Intercept -24.6

Regression Coefficient 1.05

Standard Error 0.35

Correlation .80

Analysis of Variance for a 4th Degree Polynomial

Source	d.f.	Mean Sq.	F
Linear	1	144.09	61.50
Quadratic	1	13.36	5.70
Cubic	1	62.19	26.54
Quartic	1	.81	
Deviation	2	2.34	
TOTAL	6	(Sum of Squares = 225.14)	

2. Compute critical difficulties for each subject as detailed in Stages I and II (11).

3. Discard those subjects whose critical difficulty is smaller than the largest difficulty value for the reference problems.

4. For the retained subjects and for all items, there will be a set of times to correct solution. From each of these times, subtract the speed constant for that subject. There will result a matrix of reduced times, with a row for each subject and a column for each item, but with empty cells for those items failed or abandoned.

5. The D_{Tr} value for each problem is then computed by averaging solution times for each problem across all subjects (i.e. find the mean for each column).

6. At this point it is possible to compare the resulting D_{Tr} values with the critical difficulty for each subject. For some subjects and some problems the D_{Tr} values will be greater than the lowest c D_{Tr} values. It therefore becomes necessary to remove the speed adjusted solution times from the matrix and to recompute new D_{Tr} values. This is an iterative process in which subject response times to particular items will be included or excluded at different times until 'an acceptable' state is reached. What is meant by 'acceptable' is problematic. Both Furneaux and Brierley agree that the terminal point for Stage III has to be decided by subjectively ascribing D_{Tr} values to some of the items.

The Present Study : Stage III:

In introducing the Stage III procedures, Furneaux states:

> "For every subject in the experimental group a value of a K_{Tr} can now be computed, using his O_r responses to the reference problems according to the method described in Stage I. His value of c D_{tr} is then calculated, again using the methods of Stage I".

As pointed out above (Stage III 1), this prescription is ambiguous. Brierley's interpretation of this procedure requires the use of item difficulty values from Group M to obtain speed constants for his subjects. This is the procedure adopted here. Steps 1 to 4 of this stage are then readily completed. However, as Furneaux notes, for nearly all subjects, some of the speed adjusted correct times were "relevant to I_a (t_r) inputs. A series of interations are thus necessary to remove such data and their effects from the analysis".

TABLE 20. Item difficulties on the Mill Hill following 10 iterations.

STABLE ITEMS, ITERATION 1.

Item No.	No. of Subjects
1	43
2	43
3	42
4	43
5	38
6	43
7	43
8	43
9	42
10	40
12	40
13	43
15	32
17	8
19	10
23	10
24	1
25	10
29	6
30	2
31	1

TABLE 20. Continued.

ITEMS WITH MULTIPLE DIFFICULTY VALUES.

Item No.	Range*	No. of Subjects
11	8888	39
	9013	41
14	10438	14
	10796	18
	10393	16
	10958	19
16	10768	14
	11020	16
18	10249	19
	10494	16
	9977	14
20	11115	17
	9857	8
21	9965	11
	9701	12
	10080	13
22	10097	16
	9956	15
28	10190	5
	11260	7

* Range x 10000

A computer programme to undertake these iterations was developed and used on the relevant data. A series of 10 iterations produced some interesting observations, not reported by Furneaux or Brierley, presumably because more than 1 or 2 iterations would be exceedingly laborious without a computer. The use of the computer in the present study in fact highlighted a problem encountered by both these investigators, the solution of which brings into question the scaling procedures.

The ensuing discussion refers to Table 20. In this Table, the 'No. of Subjects' column refers to the number of subjects whose solution times contributed to the item difficulty index. 'Range' refers to the range of difficulties generated during the 10 interations.

21 items showed stable difficulty values from the first few iterations onwards. 8 items showed recurrent levels of difficulty, depending on the number of iterations. Item 14 for example returned a difficulty value of 1.0438 every fourth iteration.

4 items (26, 27, 32, 33) showed a different pattern in the analysis: after only 1 or 2 iterations, it was no longer possible to obtain difficulty values for them. This problem arises as a consequence of the method used. If the mean correct solution time for an item is higher than the highest critical difficulty obtained by any subject who solved the item, then no subjects remain for the analysis. Hence these difficulty values, for this group of subjects, are indeterminate.

In commenting on this stage of the procedure, Furneaux states:

> "At the conclusion of this process a D value of probably moderate accuracy will have been assigned to the majority of the eighty items. Some, however, will have D values greater than the values of c D_{tr} relevant to even the fastest and most continuant of subjects, and thus can only be calibrated as having D_{tr} values greater than such and such. For a similar reason, others will have to be calibrated in terms of only a very few data, and thus inaccurately" (p. 181).

Brierley, confronted with such guidance admits that

> "The precise difficulty estimate depends because of these problems on the course of the successive approximations to a small degree. In these circumstances, a degree of judgement was allowed as to whether these awkward values were allowed or not". (p. 198).

From these comments, it is apparent that the central problem at this point is the absence of any rule for allocating difficulty values to the items with recurrent and indeterminate difficulties. Once subjectivity is allowed to operate, it becomes possible to assign difficulty values which will facilitate the finding of a solution which conforms to the investigator's expectations. Thus, the arbitrariness at this point must call into question the final outcome, given that these D values are used as the basis for computing the final D values, and despite the fact that Furneaux describes them as "of probably moderate accuracy".

The use of an extended number of iterations demonstrates that even apparently successful D value determination at this stage may be misleading. What value is observed will depend on how many iterations are undertaken. Item 20 (in Table 20) could have been assigned a D value of .98 or 1.11, or 98 and 111 if, as both Furneaux and Brierley did, the D values are multiplied by 100.

In the present study, it was decided to use the average of the recurrent D values as the D_{Tr} for the item. This procedure was adopted for the 8 items with recurrent values. The remaining 4 'indeterminate' items were allocated the difficulty values obtained for them in the initial phase of iteration.

All D_{Tr} data are presented in Table 23.

<u>Furneaux : Stage IV</u>:

1. Find the largest abandonment time for each subject.
2. From this value, subtract the subject's speed constant. The result is the critical level for abandonment (i.e. unambiguous abandonment c D_{Ts}).
3. Select only abandonment times for items where D is greater than c D_{Ts}. Compute the variance of these measures (V_{Ts}). This variance should be homogeneous among subjects, given the log transformation.
4. Compute the mean ($\bar{T}_s$) of these abandonments for each subject and subtract this from each abandonment time. This difference is symbolized as $\dot{T}_s$.
5. Compute the mean $\dot{T}_s$ for each item across subjects, giving $\dot{T}_s$ (Av.).
6. $\dot{T}_s$ (Av) should not correlate with D_{Tr}.
7. Pool the $\dot{T}_s$ values for all subjects. The resulting distribution should be normal. For each subject next compute the difference $\bar{T}_s - \sqrt{2 V_{Ts}}$ at all values of D_{Tr}. This is the effective minimum of T_s.
8. This value of T_s is then used in the computation of c D_{Tr} described in Stages I and II (11). Again, an iterative procedure is adopted. This iterative process should produce a larger number of I_u (T_r) for each subject; D_{Tr} values for some problems should change and K_{Tr} for all subjects should change.

The Present Study : Stage IV:

In the implementation of Stage IV procedures, it was necessary to first exclude all subjects who abandoned 1 or fewer items. This was necessary because abandonment time variances have to be computed. As a result, this stage was initiated using the data from 83 subjects. A further 50 subjects were then excluded as a result of step 2 which requires the retention of only those subjects whose abandonments were unambiguous (i.e. beyond the maximum item difficulty). It is important to note here that the substantial loss of data may be partly due to the uncertainty of some item difficulties computed in the previous stage. That is, because some item difficulties are estimated subjectively, more subjects will be lost if the item is given a high difficulty value than if the item is given a lower difficulty value.

Step 3 also requires that the variance in the abandonment time data of the remaining subjects should be homogeneous. Furneaux tested this using Bartlett's test for homogeneity of variance and only satisfied this criterion when his data were transformed to logs. In carrying out this test, the data from each subject are treated as independent samples. Each unambiguous abandonment is taken as a unit in the sample. Bartlett's test is then used to assess the homogeneity of the variances of abandonment times among the subjects.

Bartlett's test was applied to the subjects remaining in the present analysis, using the formulae in McNemar (1969, p.285). The resulting X^2 of 25.98 is not significant (32 d f). (X^2 (Chi-square) for 32 d f computed from Winer (1970, p.665)). The meaningfulness of this test on the present data is however highly questionable. 13 of the 33 variances were based on data from 2 items. While McNemar (1969) does not comment on size of the samples used in Bartlett's test, Winer (1970 p.95) states that this test should not be used when the number of units in any sample is below 3 and most should be more than 5. Only 3 subjects satisfied the latter criterion. The unsatisfactoriness of the data is due partly to the uncertain difficulty estimates, as noted above, and also partly to the few abandonments.(This latter point is discussed in a later chapter).

A further requirement at this point is that item difficulties and the $\dot{T}_S$ (Av) values for the items should not be correlated. Data for this test are presented in Table 21, from which it can be seen that only 10 data sets were available. The rank correlation for these values is - .02 which is clearly consistent with Furneaux's specification. For his own data, Furneaux found a correlation 0.11. It is possible to infer from the d.f. in his significance test (i.e. 31) that 32 items were used. This suggests that he too lost a substantial amount of data at this point, although he did not comment on this.

TABLE 21. $\dot{T}s(Av)$ for items and their D_{Tr} values.

Item No.	No. of Subjects	$\dot{T}s(Av)$*	$D_{Tr.}$*
17	1	-3662	11992
20	1	0545	10486
24	14	0429	14417
26	4	-1427	12037
27	23	0281	14847
29	4	0331	11316
30	1	-0821	9858
31	30	-0604	14374
32	21	0484	13616
33	2	1919	11649

* Data x 10000

Steps 8 and 9 at this stage require the computation of the lower bound of the abandonment time distribution for each subject. With this value available, it is possible to recompute individual critical difficulties and hence refined estimates of item difficulties. The present attempts to accomplish this proved unsuccessful. Firstly, subjects with fewer than 2 abandonments had to be excluded. Secondly, only those abandonment times, corrected for speed and above the critical abandonment level (c D_{T_S}) (i.e. unambiguous abandonments) for each subject can be employed. When the remaining data were subjected to these exclusion criteria, only 35 subjects provided usable data. Of these, only 5 had c D_{T_S} measures which were larger than the difficulty level for the most difficult of the reference items. It would obviously not be meaningful to obtain difficulty measures for the items using data for so few subjects. Any further attempts at scaling had therefore to be abandoned.

This somewhat abrupt ending was anticipated to some extent because of the substantial and progressive data loss entailed in the scaling procedures. Given that the programmes used to compute the various measures were thoroughly checked, it is doubtful if the reason for termination resides in computation errors. Other more plausible explanations are possible. The first is the high data loss as a consequence of a substantial proportion of subjects abandoning 1 or 0 items. This may in turn be a consequence of the test situation. Both these points are discussed more fully in a later chapter. A further reason, as suggested earlier, is the uncertainty of the item difficulty estimates for a substantial number of the items. Were the present outcome an isolated instance, then doubt would be cast on the entire scaling procedure used in this study. The ambiguities in Furneaux's presentation could have been reinterpreted and the scaling would then have to be repeated. However, Brierley (1969) had to terminate his analysis at the same point and for the same reasons as the present writer. It will also be seen later that scaling of the Matrices [redacted] had to be abandoned at the same point, again for the same reasons. The outcome of scaling for the Mill Hill is therefore unlikely to be simply a chance occurrence.

Despite the premature ending of the scaling exercise, it was still possible to compute more refined estimates of the speed constants for all subjects, using the difficulties for the reference items which were obtained at the end of Stage III, Step 6. These refined estimates are presented in Table 22.

With the somewhat inadequate data available at this point, it is still possible as well to undertake a "crucial" test of Furneaux's model, viz., the relationship between item difficulty and correct solution time, specified in the following equation:

$${}_{a}\bar{T}_{r} = {}_{M}D_{T_r} + {}_{a}K_{T_r}$$

TABLE 22. Speed constants for subjects on the Mill Hill and Matrices. Data x 100, corrected to the nearest whole number.

Sub	MH	MT	Sub	MH	MT	Sub	MH	MT	Sub	MH	MT
1	-25	-19	34	-24	-31	66	-17	-09	96	-28	-03
2	-20	-21	35	-17	-37	67	-32	-05	97	-22	3
3	-25	-20	36	-22	-01	69	5	-14	98	-08	-24
4	-26	-27	37	-28	-20	70	-11	-04	99	-22	-18
5	18	16	38	-12	09	71	-25	-25	100	-01	-17
6	-22	-14	39	-32	-28	72	-10	-03	101	-19	-07
9	-22	-20	40	-21	-22	73	-15	-30	102	12	26
10	-37	-18	41	-01	13	74	-28	-30	103	07	-05
11	-09	-10	42	-22	-27	75	-25	-11	104	-02	-29
12	-19	-18	43	-22	-23	76	01	-25	105	-16	-11
13	-35	-31	44	-09	-09	77	11	10	106	-08	-10
14	-15	-10	45	-24	-22	78	04	18	107	-08	10
15	-19	-13	46	-05	18	79	13	-20	108	09	13
16	-38	-20	47	17	41	80	-26	-20	109	02	07
17	-10	09	48	-38	-46	81	-18	-03	110	-06	-08
18	-11	-12	49	-33	-21	82	-22	04	111	04	-03
19	-28	-06	50	-08	-07	83	08	06	113	-35	-45
20	07	20	51	-07	-13	84	-19	-29	114	-23	-09
22	-11	-20	53	-25	-10	85	-03	-13	115	-12	-17
23	-13	-23	54	-01	-22	86	-16	-19	116	-27	-29
25	-06	11	56	-30	07	87	-08	05	117	-27	-09
26	-23	06	57	-31	-21	88	-20	-10	118	-29	-23
27	-21	-14	58	-28	-13	89	-09	-05	119	-20	-21
28	-14	-17	59	-09	-07	90	-27	-16			
29	08	00	60	-16	-26	91	-13	-07			
30	-04	-14	61	-21	-18	92	-22	-03			
31	-18	-10	62	-32	-16	93	-14	08			
32	-24	-27	64	-11	-05	94	-06	03			
33	-30	-23	65	23	-22	95	-29	-15			

Obviously, the adequacy of this test is a function of how accurately the various parameters have been estimated. For the Mill Hill data, the main limitations have already been cited. The speed constants are probably more adequate than the estimates of item difficulties although neither are fully acceptable.

The procedure for conducting this test requires the subjects to be divided into 4 groups on the basis of their speed constants, the groups ranging from fastest to slowest. For each group on each item, $\dot{T}_r$ (Av) values are computed and plotted against item difficulty. If the model holds for the data, then, as Furneaux states

> "For each sub-group separately the slope of the best fitting straight line should have a value of 1.0".

As was noted earlier Furneaux did not conduct any formal test of this relationship. He simply relied on an inspection of the data.

The data for this test are presented in Table 23.

Some preliminary comments are necessary before presenting the results of these tests.

The data from the 110 subjects were divided into 4 sub-groups (1, 2, 3, 4, fast to slow) on the basis of their speed measures. Two subjects were randomly excluded to enable equal numbers in each group. This in turn facilitated the computations. Thus, each group consisted of 27 subjects. $\dot{T}_r$ (Av) was then computed separately for each sub-group on each item, as shown in Table 23. This table also contains the data for all subjects.

The next task was to select the item difficulties. This was problematic because of the uncertain value of difficulty for some items and the indeterminate nature of the difficulties for the remainder. Because of these factors, it was decided to eliminate all such items in carrying out the "crucial" test. One further item (No. 23) was excluded because it's $\dot{T}_r$ (Av) value was based on data from only 4 subjects in group 1. The items excluded are marked Y in Table 23. 15 items remained for the final analysis.

The determination of the slope of the best-fit line to the data also posed some problems. Firstly, the data to be used do not satisfy all the assumptions necessary for computing regression coefficients. Secondly, there is no basis for deciding which of the two variables is predictor and which criterion. There are two regression lines and no basis for deciding which is to be used. This problem can be overcome by using the correlation as the index of slope if the two variables are standardized to have unit variance.

TABLE 23. Mill Hill items. $\dot{T}_r$(Av) values for all subjects and for the 4 sub-groups. Data X 100, correct to 0 decimal places. Log times.

GROUP			All		1		2		3		4	
ITEM	D	N_D	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N
1	88	43	89	109	91	26	90	27	86	27	90	27
2	73	43	73	109	77	26	74	27	73	27	67	27
3	79	42	79	106	75	26	80	25	80	26	78	27
4	74	43	74	110	75	27	74	27	74	27	74	27
5	86	38	87	102	85	25	84	26	88	26	89	23
6	84	43	84	110	84	27	85	27	82	27	85	27
7	80	43	79	107	82	26	82	25	76	27	74	27
8	79	43	78	110	75	27	76	27	80	27	79	27
9	79	42	79	108	81	26	81	27	75	26	79	27
10	83	40	80	101	80	25	86	25	77	25	79	25
11y	90	40	87	109	89	27	93	26	85	27	84	27
12	81	40	79	104	80	25	80	26	76	26	79	25
13	85	43	80	109	84	26	85	27	76	27	73	27
14y	106	16	97	94	96	19	100	26	91	25	101	23
15	91	32	90	85	97	14	92	22	85	22	88	25
16y	109	15	103	103	106	23	101	26	106	26	100	27
17y	120	8	104	58	91	11	111	12	109	17	101	17
18y	102	16	95	80	93	16	96	21	95	19	96	23
19	110	10	104	69	111	11	104	18	101	22	103	17
20y	105	12	99	60	104	10	99	16	102	18	90	15
21y	99	12	94	58	103	10	96	16	92	15	88	16
22y	100	16	95	61	92	10	95	16	97	15	98	19
23y	97	10	97	34	119	4	93	9	82	8	103	13
24y	144	1	113	39	114	4	118	13	100	8	114	13
25	101	10	99	44	102	7	99	13	101	13	95	11
26y	120	14	110	40	103	2	111	9	108	12	111	17
27y	148	5	118	26	132	6	96	5	116	5	121	9
28y	107	6	111	34	128	9	111	7	98	7	106	11
29y	113	6	105	51	117	6	100	14	107	14	102	16
30y	99	2	108	17	89	1	112	4	111	6	106	6
31y	144	1	125	17	131	4	122	3	112	5	135	5
32y	136	15	122	37	113	7	136	7	123	14	116	9
33y	116	9	112	29	102	5	127	10	103	7	107	7

TABLE 24. Regression analyses on difficulty and solution times for the Mill Hill data. Based on difficulty estimates for 15 items. Tabled data corrected to 2 decimal places.

		GROUP			
	All	1	2	3	4
Regression Coefficient	0.87	1.02	0.84	0.84	0.86
S.E. Regr.	0.04	0.08	0.05	0.08	0.11
Intercept	9.55	-1.43	13.49	10.60	9.09
Correlation	0.98	0.96	0.97	0.94	0.91
Z	3.25	0.25	3.2	2.0	1.27

Z values for regression coefficients computed as:-

$$Z = \frac{1\text{-Reg. Coeff.}}{\text{S.E. Regr.}}$$

Z values greater than 1.96 are significant at the 5% level (2-tail).

TABLE 25. Mean squares for a 3rd degree orthogonal polynomial analysis of variance on Mill Hill data for all subjects and the four sub-groups. Figures in brackets are the degrees of freedom.

		GROUP			
	All	1	2	3	4
Linear (1)	1054.24	1453.03	979.17	948.52	1030.43
Quadratic (1)	3.31	.64	13.43	.06	9.40
Cubic (1)	6.74	31.65	1.41	31.33	2.10
Deviation (11)	2.20	7.29	3.55	8.99	18.17

The usual equation for the slope coefficient is given as

$$b = r_{xy} \frac{Sx}{Sy}$$

where r_{xy} is the correlation

S_x and S_y are the variances of x and y.

If $S_x = S_y = 1$ (standard form), then

$b = r_{xy}$ (see McNemar 1969, p. 133-137)

Finally, there is the problem of the linearity of the best fit line. The regression has to be shown to be predominantly linear, particularly if r is to be used as an index of slope. r itself is a limited index however as its magnitude is inflated by the part-whole nature of the data.

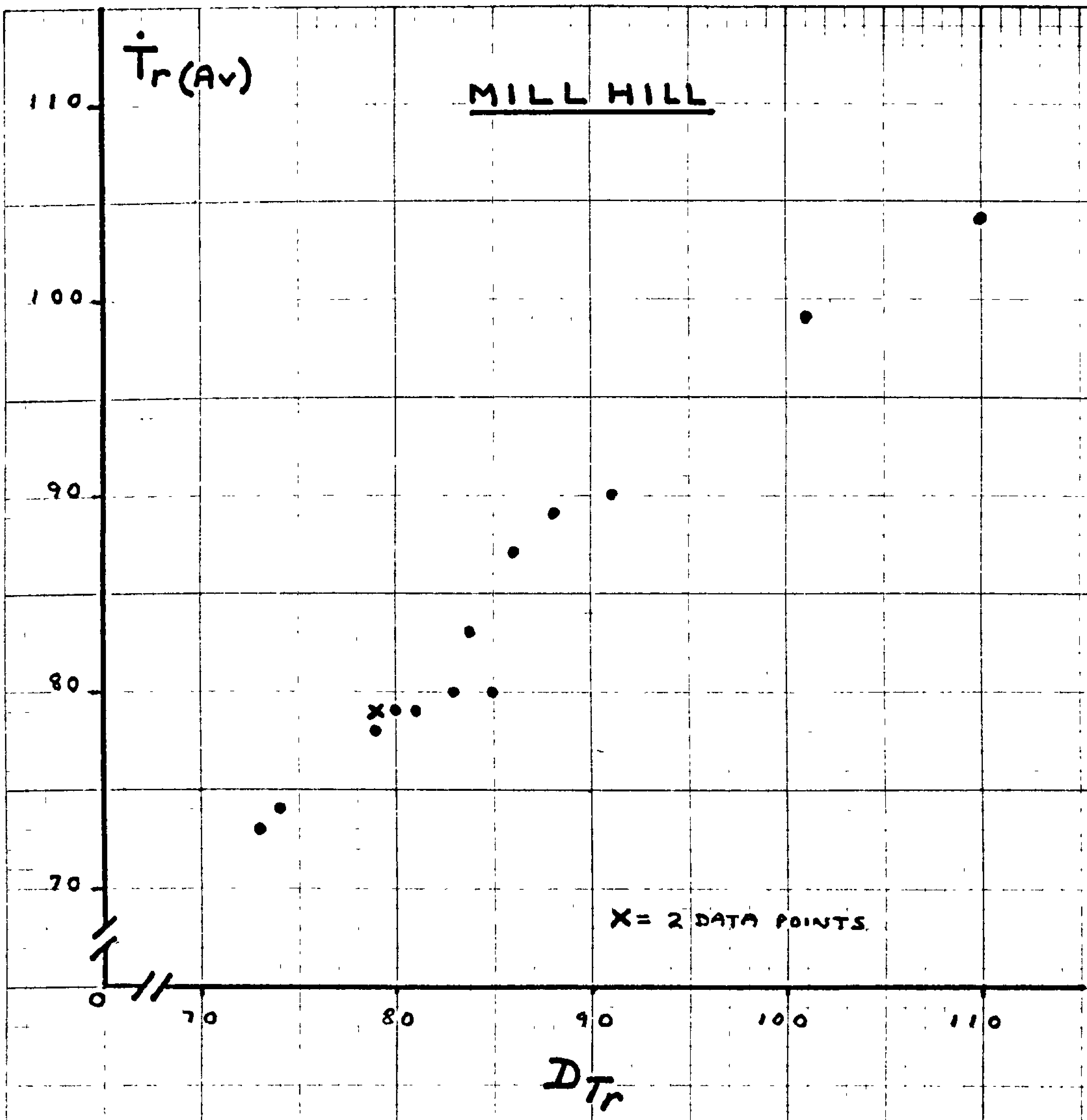
For the present data, the relationship between difficulty and $\dot{T}_r$ (Av) was examined using the B.M.D. 05R polynomial regression programme (Dixon 1970). It is used here as an indicator of the relationships, rather than as a valid statistical procedure for testing the viability of Furneaux's procedures. Item difficulty was used as the predictor in these analyses. The data for all subjects are plotted in Fig. 31 and for the four sub-groups in Fig.32.

The results of the regression analyses are presented in Table 24. Also included in this table are the Z values for the test of departure from unit slope.

From the data in this Table, it would appear that the departures from unit slope are not significant for groups 1 and 4, whereas those for all subjects and groups 2 and 3 are significant. In terms of Furneaux's model the data from group 1 and 4 appear to support his assertions about the relationship between difficulty and solution time. These data are based only on the coefficients for one of the two possible regression analyses.

The correlations in this Table suggest that the data for all groups show only trivial departures from unit slope and indicate that if these data are transferred to standard scores, the outcome would be closer to Furneaux's criterion. However, these correlations are somewhat inflated by part-whole effects.

A question not considered by Furneaux is the adequacy of the linear fit to the data. The use of a polynomial regression procedure enables an approximate assessment to be made. The index used is the mean square derived from the analysis of variance. These data were obtained using the BMD 05R polynomial regression programme. The outcome for all subjects and each of the four groups is presented in Table 25. From this Table it can be seen that in each case, the most substantial term is the linear. Higher orders of curvature account for very little of the variation. Whereas the


MILL HILL
$\dot{T}_r$ (Av)
110
100
90
80
70
0
70
80
90
100
110
D_{Tr}
X = 2 DATA POINTS

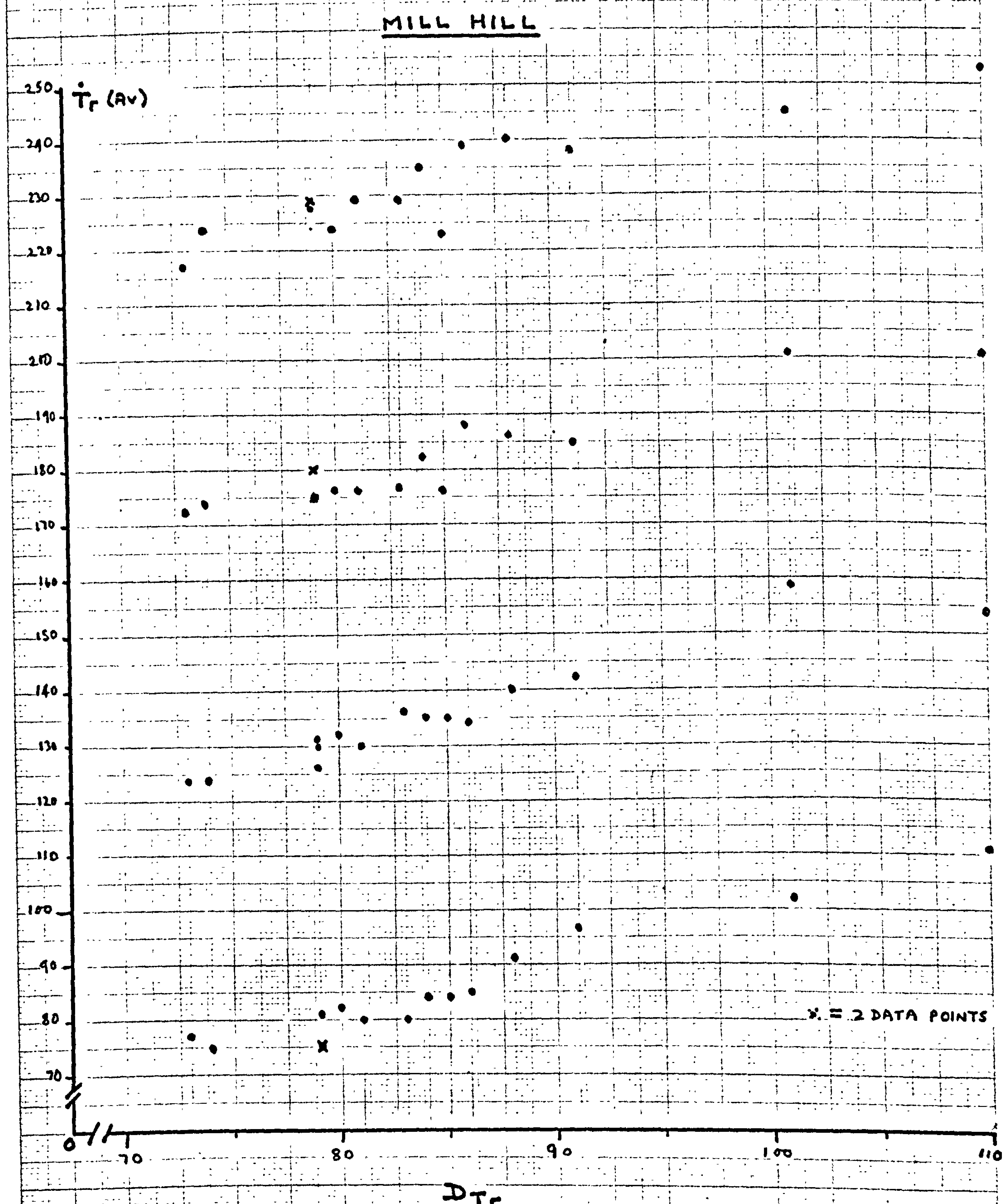

MILL HILL
$\dot{T}_r$ (Av)
250
240
230
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
0
70
80
90
100
110
D_{T_r}
x = 2 DATA POINTS

TABLE 26. Time difficulties for reference items for Groups M and F on the Matrices. Based on 12 and 15 subjects, respectively.

Item	$D_{Tr}(M)$	$D_{Tr}(F)$
2	.8363	.6897
3	.6809	.4912
4	.9906	.7210
5	.7608	.6250
6	1.0026	.8206
7	.9413	.7518
9	1.0699	.8487
21	1.1594	.8685
25	1.2750	1.0266
26	1.5090	1.2839

TABLE 27. Critical difficulties for Groups M and F at the end of Stage II, Matrices.

M		F	
Sub.	$_{c}D_{Tr}$	Sub.	$_{c}D_{Tr}$
88	1.6928	67	2.4006
14	1.7.94	104	2.4006
31	1.8161	60	1.9268
19	2.1807	33	1.8261
89	1.7662	23	1.7658
70	2.0992	40	1.6176
81	1.7060	37	1.8707
72	1.6054	86	1.4553
26	1.6354	28	1.8033
109	2.1095	100	1.4522
25	1.8682	62	1.6297
102	1.5876	95	1.8142
		51	1.9434
		58	1.7645
		1o5	1.3402

TABLE 28. Results of a 4th degree polynomial regression analysis on item difficulties for Groups M and F on Matrices. Item difficulties x 100.

Intercept	-8.49
Regr. Coeffic.	0.88
Std. Err.	0.35
Correlation	0.98

Analysis of Variance for a 4th Degree Polynomial

Source	d.f.	Mean Sq.	F
Linear	1	4232.32	359.56
Quadratic	1	27.02	2.30
Cubic	1	37.87	8.22
Quartic	1	24.88	2.11
Deviation	5	11.77	
TOTAL	9		

TABLE 29. Item difficulties on the Matrices following 10 iterations.

ITERATION 1, STABLE ITEMS

Item	No. of Subjects.
1	35
2	35
3	35
4	35
5	34
6	35
7	30
8	34
9	25
12	1
13	33
14	34
15	30
17	34
18	32
19	16
21	35
22	34
23	21
25	34
26	35
27	8
28	1
30	11
31	8
33	35
34	35
35	35
36	35
37	35
38	34

ITEMS WITH MULTIPLE DIFFICULTY VALUES

Item	Range	Subs.
10	1.6403	27
	1.6315	25
11	1.8510	14
	1.8402	13
16	2.1027	5
	2.0330	3
29	1.5098	29
	1.5192	30

F-ratios for the linear components for all groups would be highly significant, none of the other ratios would be so. It would however be inappropriate to conduct formal statistical tests on these data.

The conclusions from the foregoing analyses are presented at the end of the next chapter.

4. Item Difficulty : Advanced Progressive Matrices:

The procedures for scaling the Matrices items were identical to those adopted for the Mill Hill. Once the basic set of computer programmes is available, only minor modifications are needed to accommodate the Matrices data. This commentary will present the salient features of the Matrices analysis.

63 subjects abandoned 2 or more items and the data from the 31 most continuant of these were selected for the preliminary analysis of the response items.

The mean abandonment time on the Matrices was 1.8799 log secs. (On the Mill Hill this was 1.0739 log secs.). In 'real' time, the respective means were 75.74 and 11.83 on the Matrices and Mill Hill. The distribution of abandonments is presented in Fig. 25.

For the Matrices data, it was eventually possible to select 10 reference items (No's. 2, 3, 4, 5, 6, 7, 9, 21, 25, 26) which approximated Furneaux's criteria. The difficulty values for these items for groups M and F are presented in Table 26. They were obtained from the data for 12 subjects, after 6 attempts. Group F data are based on 15 subjects. Critical difficulties for the subjects in these groups are presented in Table 27.

From these Tables it can be seen that the greatest difficulty for a Group M reference item was 1.5090 and the lowest critical difficulty for any subject was 1.5876. For Group F, the respective values were 1.2839 and 1.3402. These data satisfy step 12A of Stages I and II.

The regression analysis produced a linear coefficient of .88 (s.e. .05). From Table 28 which summarizes the analysis, it can be seen that only the linear component is substantial.

The iterations of the Stage III analysis also produced three sets of item difficulties. The outcome of this stage is summarized in Table 29. 31 items showed stable difficulty values, 4 varied as a function of the iteration process and 3 items were of indeterminate difficulty.

The test for homogeneity of variance was carried out on the data for 35 subjects whose abandonments were beyond the maximum item difficulty. A X^2 of 51.41 (34 d.f.) was found significant at p less than .05, indicating that these variances are not homogeneous. Again, the meaningfulness

TABLE 30. $\dot{T}s(Av)$ values for items and their D_{Tr} values. Matrices.

Item	No. Subjects	$\dot{T}s(Av)$	D_{Tr}
11	11	-.0271	1.8456
12	17	-.0705	2.5559
16	6	.0835	2.0679
19	22	.2495	1.7937
20	14	.0041	2.0389
23	22	-.0385	1.6708
24	19	.0220	2.3308
28	11	.0898	2.3722
30	2	-.1151	1.9409
31	6	-.1126	1.9035
32	23	-.0110	2.2323

TABLE 31. Matrices items. $\dot{T}_r$(Av) values for all subjects and for the 4 sub-groups. Data X 100, corrected to 0 decimal places. Log times.

GROUP			All		1		2		3		4	
ITEM	D	N_D	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N	$\dot{T}_r$(Av)	N
1	105	35	104	108	102	27	109	25	106	27	103	27
2	86	35	88	110	87	27	90	27	89	27	88	27
3	69	35	69	110	67	27	72	27	69	27	69	27
4	97	35	97	102	97	25	100	27	97	24	99	25
5	81	34	80	110	81	27	77	27	81	27	81	27
6	102	35	100	105	102	25	101	25	101	26	98	27
7	95	30	96	109	96	26	98	27	95	27	97	27
8	134	34	132	82	138	19	131	22	132	18	130	22
9	107	25	108	103	114	25	105	26	111	25	105	25
10y	164	26	161	94	161	19	158	26	166	24	163	23
11y	185	13	173	71	169	20	172	19	177	16	182	14
12y	256	1	206	34	198	9	212	13	193	5	223	7
13	124	33	120	97	117	25	119	25	121	22	126	24
14	139	34	143	104	136	26	144	26	142	24	151	27
15	150	30	148	79	146	19	144	21	156	18	151	20
16y	207	4	206	45	200	8	198	15	215	11	220	10
17	110	34	108	105	109	25	105	27	108	25	111	27
18	144	32	142	100	144	27	136	26	139	22	151	25
19	179	16	173	69	168	17	175	21	173	14	179	16
20y	204	2	207	21	215	5	217	4	191	6	209	6
21	111	35	117	109	115	27	110	27	111	27	112	26
22	137	34	136	103	136	25	129	26	137	24	143	26
23	167	21	168	70	170	18	159	18	169	16	172	17
24y	233	3	208	44	193	9	203	9	205	11	225	14
25	125	34	120	103	121	26	120	26	118	23	121	26
26	149	35	148	106	144	26	149	27	151	24	152	27
27y	191	8	182	66	185	15	179	19	170	15	195	17
28y	237	1	198	25	174	5	199	4	201	6	207	9
29y	151	29	150	91	152	22	145	24	147	21	158	24
30	194	11	177	88	163	22	180	22	175	21	194	21
31y	190	8	189	49	183	13	189	13	186	9	199	14
32y	223	1	225	6	205	1	223	1	234	1	230	3
33	103	35	102	108	98	26	105	27	104	26	106	27
34	105	35	104	107	104	26	102	27	107	27	108	26
35	98	35	97	110	99	27	98	27	98	27	98	27
36	81	35	82	110	83	27	78	27	89	27	83	27
37	121	35	119	105	115	26	113	26	127	27	123	24
38	106	34	109	104	111	23	109	27	114	26	107	26

TABLE 32. Regression analysis on difficulty and solution times for Matrices data. Based on difficulty estimates for 27 items. Tabled data corrected to 2 decimal places.

	GROUP				
	All	1	2	3	4
Regr. Coeff.	.93	.87	.90	.92	1.02
S.E. Regr.	.02	.04	.02	.03	.02
Intercept	7.29	13.87	9.58	9.99	-1.13
Correlation	.99	.98	.99	.98	.99
Z	3.5	3.25	5.0	2.67	1.00

Z greater than 1.96 significant at the 5% level (2-tail)

TABLE 33. Mean squares for a 2 degree polynomial analysis of variance on the Matrices data for all subjects and the four speed sub-groups. Figures in brackets are the d.f.

	GROUP				
	All	1	2	3	4
Linear (1)	21018.99	18359.46	19821.93	20631.41	25513.66
Quadratic (1)	66.08	287.46	6.02	136.97	3.58
DEVIATION (24)	7.04	21.87	13.52	21.98	11.28

of such an analysis on these data is questionable, given that 16 of the 35 variances were based on 2 items and a further seven on 3 items (Winer 1970). The X^2 for 34 d.f. was computed from the formula given in Winer (1970, p. 665).

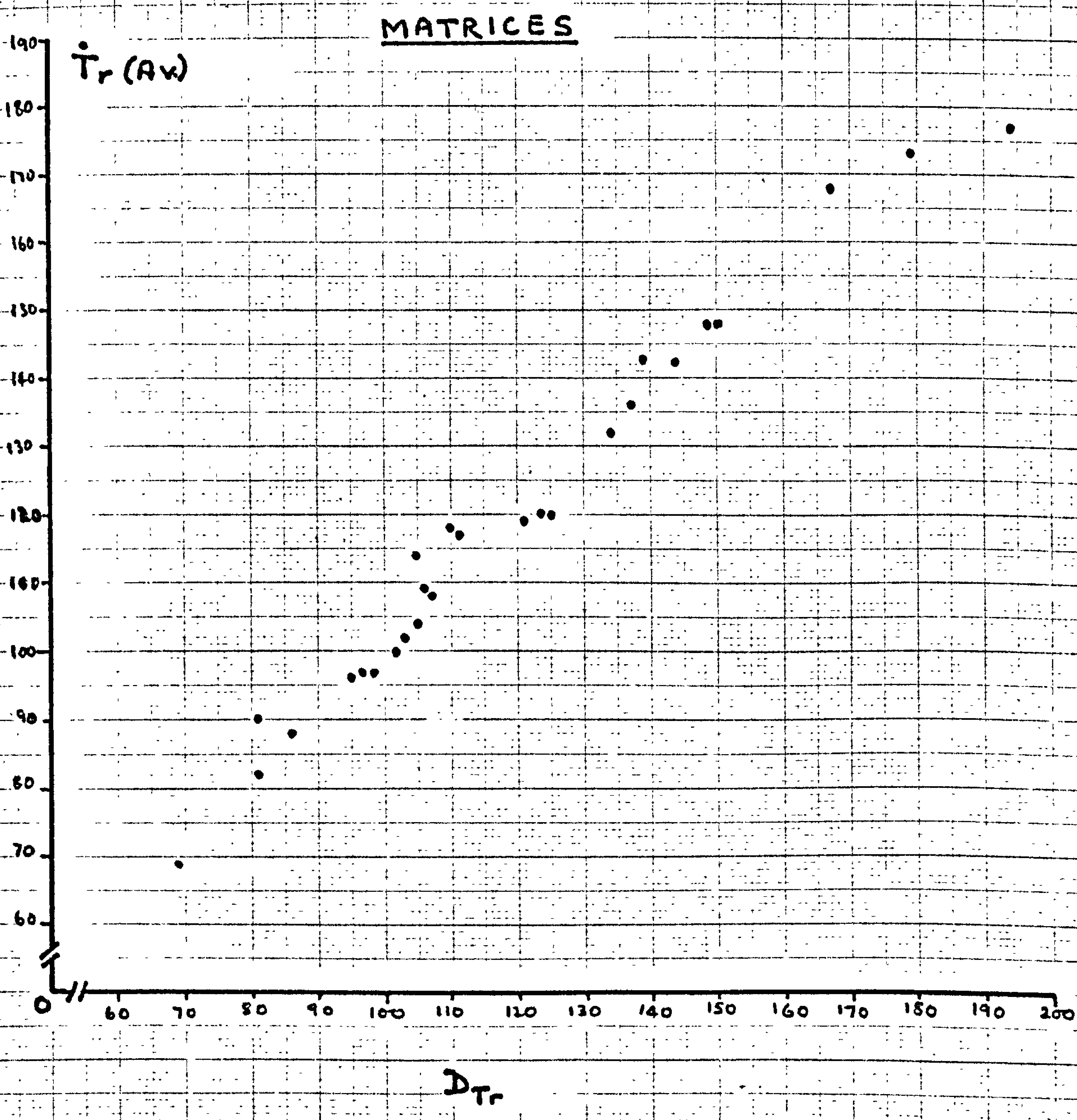
For the Matrices items, it was also found that difficulty and $\dot{T}_s$ (Av) were uncorrelated. However, because of the comparatively few abandonments only 11 items contributed to this analysis, producing a correlation of -.06. When those items based on the data for two or fewer subjects are excluded, the resulting correlation is approx .2. Again, the meaningfulness of the coefficient is questionable, being based on data for 8 subjects. These analyses are nevertheless consistent with Furneaux's specifications. The items entered into this analysis, together with the $\dot{T}_s$ (Av) values and item difficulties are presented in Table 30.

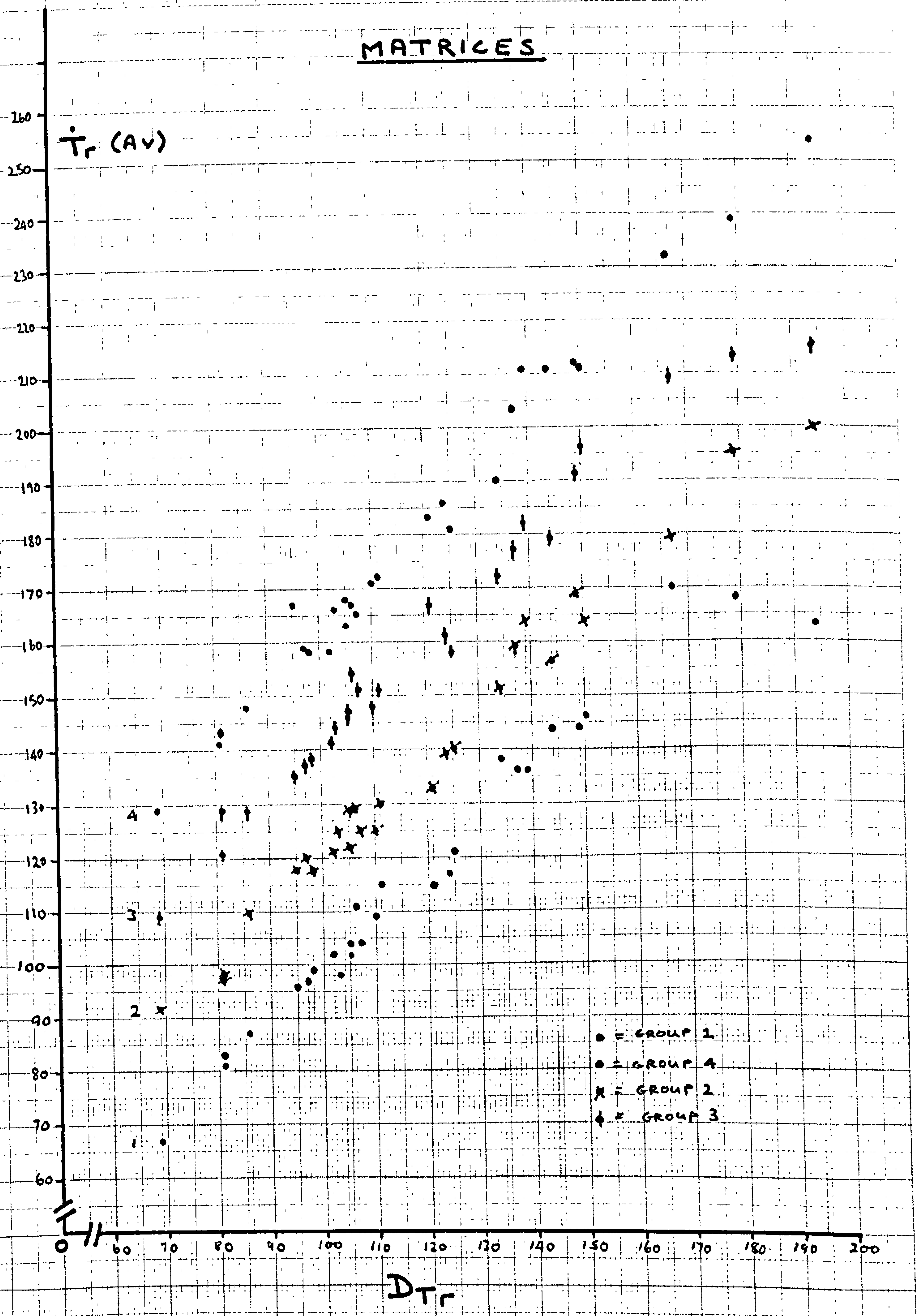
As with the Mill Hill further attempts to refine the difficulty values of the Matrices items proved unsuccessful, for the same reasons. In the present instance, only 7 subjects remained with critical difficulties greater than the maximum difficulty for a reference problem. However, it was again possible to calculate new speed constants for all subjects. These constants were presented in Table 22.

The "crucial" test on the Matrices data were carried out in the same way as for the Mill Hill. The same limitations apply to the Matrices analysis as was the case with the Mill Hill. The data for this test are presented in Table 31. All analyses were carried out on items other than those marked 'Y' in this Table. The 'Y' items are those of uncertain difficulty or those whose $\dot{T}_r$ (Av) values were based on less than 10 subjects. 27 items remained for the analysis. The outcome of the regression analysis is presented in Table 32.

The regression coefficients for all groups as well as the total sample are very close to Furneaux's criterion of unit slope. However, as suggested by the Z values reproduced in the table, all coefficients except that for group 4 show significant departures from unit slope. None of the correlations, as indicies of slope for standard scores with unit variance, show substantial departures from unity, the lowest correlation being .98.

The polynomial regression analyses show that by far the largest component is the linear. Although cubic trends were not computed, all the residual mean squares, as well as those for quadratic effects are very small compared to those for the linear component. The various mean squares are presented in Table 33.


MATRICES
$\dot{T}_r$ (Av)
190
180
170
160
150
140
130
120
110
100
90
80
70
60
0
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
D_{Tr}


MATRICES
$\dot{T}_r$ (Av)
D_{Tr}
● = GROUP 1
● = GROUP 4
✗ = GROUP 2
ɸ = GROUP 3
1
2
3
4
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
0

5. Discussion:

The present investigation represented an attempt to examine the applicability of Furneaux's scaling procedures to sets of data from two different tests. It was noted that some of the crucial empirical findings reported by Furneaux were unsupported by relevant statistical tests. In the process of conducting the present investigation, it was observed that despite their seeming rigour and detail, Furneaux's specifications for the scaling procedures contain some ambiguities. In implementing these procedures, a number of problems arose, the most important of which was undoubtedly the substantial data loss as the analysis proceeded. As a consequence of this loss, the analysis could not be taken to its conclusion. A further consequence was that certain essential statistical tests could not be applied because of the ensuing violation of the underlying assumptions. The analyses undertaken in this part of the study were undoubtedly in violation of these assumptions and it must therefore be emphasized that the conclusions must be qualified.

On the basis of the analyses reported in the last two chapters, it was found that the logarithm of solution time for correct responses is linearly related to Furneaux's operational definition of item difficulty. The slope of the line relating these two indices approximates unity when the regression coefficients are computed from the raw scores (log time). If these scores are expressed in standard form with unit variance, then, as indexed by the correlation between them, the approximation to unit slope is even better. It should also be noted that this relationship emerged even though the estimates of item difficulty were not as refined as they could have been had no appreciable data loss arisen.

The regression coefficients for the 8 sub-groups ranged from 0.84 to 1.02 on the two tests. Although only 3 of the 8 coefficients did not differ significantly from unit slope, it has to be noted that these slopes were based on a proportion of the items and that the difficulty values are not as refined as the values computed by Furneaux. With these limitations in mind, it is suggested that these values are sufficiently close to unity to provide some support for Furneaux's assertions. It could of course be argued that Furneaux did not attempt a statistical assessment and that his use of arbitrary values for difficulty could easily have shifted the slopes of his data to unity.

This, in fact, is a major problem of Furneaux's approach which has major implications for the speed/difficulty relationship. It is possible that for the lower difficulty region, the relationship with speed is linear but that for the upper difficulty region the slope is different, so that overall, the regression line is curvelinear. If the researcher <u>believes</u> that the overall slope is linear, and if difficulty values in the upper

Table 34. Number of abandonments on Mill Hill and Matrices.

SUBJECT	MH	MAT	SUBJECT	MH	MAT	SUBJECT	MH	MAT
1	0	0	47	5	1	93	7	0
2	1	0	48	1	1	94	10	3
3	7	3	49	2	0	95	3	2
4	1	2	50	7	3	96	0	6
5	17	0	51	1	4	97	4	1
6	7	4	53	3	1	98	0	0
9	5	4	54	4	2	99	1	1
10	9	1	56	4	4	100	9	3
11	14	1	57	0	1	101	6	0
12	0	0	58	1	3	102	1	2
13	1	2	59	3	4	103	3	1
14	6	5	60	6	4	104	4	2
15	0	1	61	5	1	105	12	4
16	1	3	62	3	3	106	4	2
17	0	0	64	3	2	107	4	0
18	9	1	65	2	0	108	11	5
19	4	2	66	4	1	109	2	2
20	4	3	67	2	2	110	7	1
22	7	5	69	13	1	111	19	9
23	4	2	70	11	4	113	3	0
25	10	4	71	1	4	114	2	8
26	0	4	72	0	2	115	2	1
27	4	6	73	3	4	116	2	1
28	7	3	74	7	0	117	3	2
29	4	2	75	10	1	118	11	4
30	6	4	76	5	1	119	5	1
31	1	3	77	5	1			
32	5	1	78	0	1			
33	6	2	79	12	0			
34	3	7	80	0	1			
35	7	1	81	7	5			
36	2	1	82	4	6			
37	1	3	83	14	3			
38	1	0	84	0	0			
39	0	0	85	10	2			
40	10	3	86	2	3			
41	8	4	87	2	6			
42	5	1	88	7	3			
43	5	3	89	5	3			
44	4	2	90	2	1			
45	0	4	91	2	1			
46	11	3	92	2	0			

TABLE 34a. Abandonments on the Mill Hill and Matrices

No. of Abandonments	Frequency Mill Hill	Matrices
0	14	18
1	13	29
2	13	18
3	10	18
4	14	16
5	10	4
6	5	4
7	11	1
8	1	11
9	3	1
10	5	
11	4	
12	2	
13	1	
14	2	
15	0	
16	0	
17	1	
18	0	
19	1	

TABLE 34b. Distributions of total scores at the upper end of the Mill Hill and Matrices.

Mill Hill		Matrices	
Total Score	f	Total Score	f
33	1	38	0
32	2	37	0
31	5	36	1
30	4	35	7
29	4	34	11
28	1	33	8

TABLE 35. Abandonments on the Mill Hill and Matrices. Number of abandonments on each test associated with number of abandonments on the other.

		MATRICES			
		0	1	2+	TOTALS
	0	6	4	4	14
MILL HILL	1	2	2	9	13
	2+	10	23	50	83
	TOTAL	18	29	63	110

range have to be subjectively estimated, then it is possible to chose difficulty values which give an overall linear slope, whereas in reality the slope overall is curvelinear. In view of the earlier comments of both Furneaux and Brierley, and the observations in the present study, it is suggested that as yet there are insufficient grounds for asserting that over the full range of difficulty values, the speed/difficulty relationship is linear. Further, the present evidence that the most substantial portion of variance is linear does not contradict this view. It will be recalled that the test was undertaken only on a proportion of the items and that these were in the lower difficulty range.

In the course of the analyses reported here, it became obvious that a major factor in this study was the loss of data due to insufficient abandonments. This problem is discussed in the next chapter.

6. A 'Post Hoc' Analysis of Abandonments:

a. Introduction:

The frequencies with which subjects showed different numbers of abandonments are given in Tables 34 and 34a. The distributions for these data were shown in Figs. 24 and 25.

The number of abandonments associated with each item of the Mill Hill and Matrices were given in Tables 11 and 14 and the mean and standard deviation of the log abandonment times in Tables 12, 13, 15 and 16.

On the Mill Hill and Matrices, about 24% and 42% respectively, of the subjects abandoned 1 or 0 items. As a consequence, the data available for difficulty scaling were substantially reduced. This, as well as a number of other factors led to only approximate values for item difficulty. The purpose of this chapter is to examine some of the main factors likely to have been responsible for the premature termination of the difficulty scaling. It is important to note here that in the only other study which used Furneaux's procedures, that of Brierley (1969), difficulty scaling could not be completed, probably for the same reasons as affected the present study.

In this chapter, two general problems will be considered. The first concerns the possible reasons for the reduced abandonments data and the second, the likely effects of these factors on the item difficulty estimates.

b. Factors likely to have influenced the number of abandonments:

i. One of the factors likely to account for the low level of abandonments is the 'easiness' of the items relative to the ability level of the subjects: if the majority of the items are 'easy', then there would be no need to abandon attempts at solution. An examination of the total score distributions for both tests would not support this contention. Both distributions have 'tails'. On the Mill Hill, only one subject got all answers correct whereas no subjects got all items correct on the Matrices. The 'tails' of both distributions are given in Table 34b.

ii. In discussing why he encountered a limited number of abandonments, Brierley (1969) suggested that with paper and pencil tests (such as those used by Furneaux), subjects may well feel that they can return to an omitted item, even though this is discounted in the instructions. This is however unlikely to be the case with Furneaux's study because his subjects had to enter a timing record after each item and this must to some extent have discouraged them from returning to abandoned items. Further, although Furneaux does not present any evidence on this point, he does state that "once an item had been abandoned it could not be attempted again". With automated presentation, as was made clear in the instructions given to subjects, a response is irrevocable. Brierley (1969) reports " -- many subjects indicated that in these circumstances it was better to press an equally available answer button which may well turn out to be the right one" (p. 201). The problem of guessing, and related to it, the multiple choice format of the tests, are implicit in these comments, and will be considered next.

It should be noted that the standard procedures (Thorndike 1971, p. 59-61) for introducing corrections for guessing in multiple choice tests are not relevant here. These procedures apply to total scores and cannot cope with individual items, the focus of this type of research.

It is apparent, with hindsight, that the instructions were not sufficiently explicit and prohibitive with regard to guessing. In Brierley's (1969) study, subjects were told

> "Try your best to work out each test but if
> you decide you can't find the answer, you
> can press the "don't know" button - - " (p. xxvi).

In the present study, the relevant section of the instructions on both tests stated:

> "If at any stage you feel uncertain about your
> answer, or if you would like to go on to the
> next problem, press the red button - - - - - -
> And remember that if you are at all uncertain
> of your answer or that you would like to leave
> the problem, press the red button - - - "

It is obvious from both excerpts that guessing was not strongly discouraged, although the present instructions appear, superficially at least, to have been more prohibitive. It is of course possible that the subjects who did not abandon items which were incorrect may have felt that the solutions offered were correct. Unfortunately, the circumstances of testing did not permit them to be questioned on this point.

In future studies of this type, it will be necessary to make the instructions much more explicit. It may also be possible to introduce a system for controlling guessing, as is done in research on signal detection thresholds. However, such a modification would need to be thoroughly tested beforehand as there may well be systematic individual differences in response to prohibitions and inducements.

One alternative that should be given serious consideration here is the introduction of some method for assessing subjective certainty that the answer is correct. It may then be possible to discriminate among correct, incorrect, abandoned and 'have a go' answers.

The type of experimental situation in both Brierley's and the present study was not amenable to either reinforcing the initial instructions or to monitoring the subject's certainty. In both investigations, subjects were seated in sound insulated test rooms and it would have been disruptive to intrude, even if such intrusions did not produce other complications such as the timing of the intrusion and eliminating it's effects from the data.

iii. The use of a multiple-choice format may have been important in producing an 'abandonments problem'. In the present study the Mill Hill provided 6 alternatives, the Matrices 8, with the respective 7th and 9th alternatives being 'abandonment'. Brierley's test offered 14 alternatives, one of which was abandonment. With the letter-series, the theoretical maximum number of alternatives is 26 but in practice, the item content serves to narrow this range considerably, depending on the problem. Given that a guess is less likely to be correct the greater the number of alternatives, the larger the number of alternatives the more likely it is that guessing will be discouraged. It is possible that because of the number of alternatives for his items, Furneaux did not encounter any substantial problems with guessing. Certainly any future study of this type should consider carefully the choice of answering procedure.

iv. The motivation of subjects to persist is an important and subtle factor in the 'abandonments problem', in that it has implications for the peculiar scaling procedure described by Furneaux.

In the description of his study, Furneaux emphasizes that high motivation is important. As he states:

> "Instructions were designed to encourage high motivation, and the evidence suggests that this was achieved. Stress was laid on the need to persevere with items found to be difficult, rather than attempt to reach the end of the test".

Parenthically it should be noted that the instructions he gave were not reported and that he did not provide any of the evidence alluded to. Given the group test situation, one might also question whether or not subjects will persist when all around them they can see others proceeding through the test. Overt timing is also unlikely to facilitate persistence.

Following Furneaux's comments, both Brierley and the present writer devised instructions to encourage persistence. Brierley's subjects were instructed thus:

> "There is no special need to hurry and you can go at the speed you fancy best but don't waste time on anything" (p. xxvi).

In the present study, subjects were told

> " -- you can take as much time as you like to find the answer. Do not feel that you need to hurry".

In considering the role of motivation, Brierley suggests that the use of an automated procedure in conjunction with instructions designed to encourage motivation to persist, led to comparatively long solution times for some subjects, thereby raising the upper critical difficulty level. Although subjects in this study did not spend as long (in the absolute sense) on items as did Brierley's subjects, at least one spent nearly 8 minutes on one item and it was not uncommon to have solution times as long as 4 minutes. Such times need of course to be considered against the more common range of solution times of a few seconds to 1 minute.

In-so-far as instructions and automated procedures are condusive to high continuance, they produce certain consequences for scaling, notably, they will tend to increase the time difficulty of the items. This in turn means that the lower bound of the unambiguous abandonments range is raised. Because of this, the number of abandonments that could have been unambiguous will be reduced. The ensuing loss of data then has the consequence of making further refinements in scaling impossible because further refinement depends crucially on the availability of a sufficient number of unambiguous abandonments.

Another consequence of the instructions, in-so-far as they are effective, will be to increase the amount of time that a subject is likely to spend on an item. Given that he follows instructions, the more likely it is that a correct solution will be forthcoming, reducing further the chance of an abandonment.

Another factor likely to be of importance in individual testing, particularly if it is automated and therefore private, is that the testee is not subjected to the implicit pressure to work quickly that is present in group testing. That is, seeing other subjects completing the test, turning pages more rapidly etc. might have diverse effects on solution and abandonment times that are not present in individual testing. The extent to which such pressures affected subjects, and hence their responses in Furneaux's study cannot be known. It is however possible that the group as opposed to private individual testing may have a differential effect on the magnitude of solution or abandonment times.

v. It is likely that in both this and Brierley's study, the number of subjects tested was a further important factor in the incomplete scaling. Given the intricate mechanics which lead to data loss during Furneaux-type scaling, it is obviously important to test a very large number of subjects. Furneaux began with 235 subjects. It is not known how many remained to contribute to the item difficulties, but he at least appeared satisfied that the numbers were sufficient to produce reliable difficulty values. The fact that he was able to test this large a number of subjects shows two important limitations of automated individual testing. These are restricted mobility and restriction to individual testing. Whereas Furneaux was not confined to one physical setting and could test groups, both Brierley and the present writer had to bring subjects to the laboratory and test them individually. Future researchers will have to balance the greater precision and flexibility of automated testing against the more economic procedure of group testing, which inevitably carries with it more uncontrollable factors. With the development of the hardware and software for remote access to a central computer, together with microcircuitry and miniature video screens, it might well be possible in the future to test more than one subject at a time, thereby overcoming some of the limitations of automated testing.

c. <u>The effects of abandonments on item difficulty</u>:

A major question which needs to be considered is the extent to which the exclusion of the 0 or 1 abandonments subjects has influenced the outcome of the analysis. (It will be recalled that such subjects were excluded from scaling exercise). Two aspects of this question will be considered here.

The first aspect concerns the extent to which the same subjects were responsible for the comparatively large numbers of 'no abandonments'. The relevant data are summarized in Table 35. From this Table it can be seen that only 6 subjects abandoned 0 items on both tests. It can also be seen that there is a low level of predictability working from either test to the other. For example, of the 13 subjects who only abandoned 1 item on the Mill Hill, 2 showed 0 abandonments, 2 showed 1 and 9 showed 2 or more Matrices abandonments. This observation is further supported by the correlation of .21 between abandonments on these two tests (p less than .05). Although the correlation is significant for the N of 110, its value as a predictor is minor.

Assuming the validity of Furneaux's analysis, there is of course no reason to expect consistency across tests, unless it has been demonstrated beforehand that items in each are of equal difficulty and that subjects are of equal speed.

The second aspect to be considered is the extent to which the exclusion of subjects with 1 or 0 abandonments affected the crucial test proposed by Furneaux. The most appropriate way to do this is to contrast subjects excluded from the analysis with those who were included in the computation of both DT_r and $\dot{T}_r$ (Av), the two variables used in the crucial test. To do this systematically would be exceedingly cumbersome. Different numbers of subjects, even those with abandonments data, contributed to the difficulty estimates for each item. Similarly, different subjects contributed to the $\dot{T}_r$ (Av) values for each item. Thus, for reasons inherent in Furneaux's scaling procedures, an exercise of this type would be complex and probably as time consuming as the scaling exercise itself. As it is doubtful that such an exercise would have contributed anything of significance in any case, this analysis was not undertaken.

d. Conclusions:

In this chapter, an attempt has been made to consider some of the reasons for the premature cessation of scaling using Furneaux's procedures. A number of factors have been discussed including the problem of guessing, the effects of motivating instructions, the use of individual testing and the numbers of subjects tested. The extent to which each contributed to the abrupt ending cannot be known: however, it is suggested that each probably had some impact. A further question considered was the predictability of abandonment from one test to the other. Although the correlation between abandonments on both tests was statistically significant, it was of such a low order as to be of very limited value. Finally, given the nature of the scaling procedures, it is very difficult to know the effects of the exclusion of subjects for different aspects of the analysis.

VIII. AN INVESTIGATION OF SPEED, PERSONALITY AND TEST PERFORMANCE:

This section of the thesis is devoted to an analysis of the data on personality and test performance.

The original intention of this part of the study was to use the data on speed, accuracy and continuance to examine personality-related functioning on tests of intelligence. As a result of the incomplete Furneaux scaling, only speed constants could be derived for each subject. The relationships between speed, several other variables to be described shortly, and personality measures are considered in this part of the thesis. These variables were derived from the procedures described earlier.

1. Data:

The original data set available from this study was reduced to the set presented in Table 36. The reasons for the reduction are given in this chapter. The distributions for the variables were given in Figs. 14 to 30. The distributions are based on 110 subjects.

Age:

This variable was transformed using the square root transformation to achieve a closer approximation to normality. This was necessitated by the procedures used in the statistical analysis (see section on Statistical Assumptions in the Data Analysis).

P-score:

For the analysis involving the use of the personality variables, the P-score was excluded. There were two main reasons for exclusion. Firstly, the items defining P in the P.E.N. were in the process of being refined and the scale was not yet standardized. Secondly, there are no specific hypotheses implicating P.

E, N, and L-scores:

Derived directly from the P.E.N.

Total Correct Mill Hill and Matrices:

Raw scores for the subjects on these tests.

The version of the Matrices used in the present study cannot of course provide an I.Q. because it's structure is unlike that of either of the original tests which provided the items. The common techniques for computing I.Q's from raw scores involve a linear transformation to a deviation quotient with an arbitrary mean and standard deviation, or by a linear transformation from the percentile equivalents of the raw scores.

TABLE 36. Final list of variables used in the present study.

VARIABLE	CODE*	TRANSFORMATION
AGE	AGE	SQ. ROOT
EXTRAVERSION	EXTVSN	-
NEUROTICISM	NEUROT	-
'LIE' SCALE	LIESCL	SQ. ROOT
TOTAL CORRECT, MILL HILL	HTOTAL	-
TOTAL CORRECT, MATRICES	MTOTAL	-
TOTAL TIME, MILL HILL	HTIME	LOG x 100
TOTAL TIME, MATRICES	MTIME	"
SPEED, MILL HILL	HSPEED	x 100
SPEED, MATRICES	MSPEED	"
ABANDONMENTS, MILL HILL	HABTS	SQ. ROOT
ABANDONMENTS, MATRICES	MABTS	"
SLOPE	SLOPE	x L100
SIMPLE REACTION TIME	1/1-RT	"
1 of 2 CHOICE TIME	1/2-RT	"
1 of 4 CHOICE TIME	1/4-RT	"
1 of 8 CHOICE TIME	1/8-RT	"
ERRORS, MILL HILL	HERROR	SQ. ROOT
ERRORS, MATRICES	MERROR	SQ. ROOT

*These codes identify the variables in the data tables.

It is the common practice for tests which span a wide age range to set up norms for each age or age-span, in effect, adjusting the scores for age differences in performance. In the present study, only raw scores have been used in the primary data. However, as will be seen later, these scores were subjected to an age adjustment through partial correlation and such adjustments produce data which in effect correspond to I.Q. A similar procedure was used for the sub-set of items that comprised the Mill Hill test used in this study.

Total Time on Mill Hill and Matrices:

The total raw time taken by each subject to complete the tests, irrespective of type of solution, was computed and the log taken. Each log score was then multiplied by 100.

Speed Mill Hill and Matrices.

KT_r from the Furneaux scaling study. Each subject's KT_r multiplied X 100.

Abandonments Mill Hill and Matrices:

Raw scores for the subjects.

Slope and Reaction Times:

Data from the pre- and post-reaction time series were combined and averaged. As the distributions of these times appear to be normal (see Figs. 27 to 30) no transformations were necessary.

To compute the slopes for the best fit straight line, a modified version of BMD 05R (Dixon 1970) was used. This computed the slope of the best fit line though 1/1, 1/2, 1/4, 1/8 means for each subject, following the procedure of Roth (1964) and Frank (1963).

The intercepts for the best fitting line were also computed for each subject but were eliminated from the final analysis because they correlated + .97 with simple reaction time.

Errors on Mill Hill and Matrices:

Raw number of errors, transformed by square root.

The final data set consisted of 19 variables.

2. Reliabilities:

The arrangements for testing, in which subjects were paid to participate in several studies apart from the present investigation, did not allow proper reliability estimates to be made for the main variables in this study. However, it is possible, from other sources, to get some idea of the level of reliability of a number of the variables.

Although there are no reliability data for the E, N, and L scales of the P.E.N., the E and N scales of the E.P.I. have re-test reliabilities of .88 and .84 after 1 year and the L scale is usually above .70 (Eysenck and Eysenck 1964).

The Mill Hill is a compound test consisting of two components, a written definition section and a multiple choice section. No reliabilities are given for the separate sections but the parallel form coefficient (Form I and II Senior) is quoted as .98 for adults (Raven 1965b).

The Matrices test used in the present study was made up of items from various versions of the Matrices but predominantly of items from sets I and II of the Advanced series in the revised version (Raven 1965a). Raven reports a retest coefficient of .91 (interval not specified) for a group of 243 adult students. While this may well be an over-estimate, it is unlikely that the level of reliability is so low as to make the use of the present test scores unacceptable.

It is not possible to know the extent of unreliability of the Furneaux speed scores.

Although the reaction time data showed high inter-correlations between pre- and post-testing (ranging from .83 to .90), these correlations can only give an approximate indication of the likely reliability of the measures used in the data analysis.

The limited information on the reliability of the variables must detract somewhat from the findings of this study but it is doubtful that the measures used here are unreliable to the extent of making this a meaningless piece of research. However, the absence of such data does mean that the various correlations cannot be corrected for attenuation. While this does not matter for prediction purposes (Mc Nemar 1969) it does have implications for some of the theoretical aspects of this study. This issue will be discussed further in relevant sections of this thesis.

3. Number of Subjects:

The Furneaux scaling study utilized data from 110 subjects. For the second part of this study, 8 subjects with high L-scores were excluded. The decision to do so was based on the supposed nature of high L, namely, that subjects with high L-scores are responding inappropriately on the P.E.N. The cut-off point was decided upon following an examination of the L-score distribution (Fig. 17). From this histogram it can be seen that the distribution drops sharply for L-scores of 7 or more. Only 8 subjects had scores greater than or equal to 7 and they were excluded. Thus data from 102 subjects are reported on here.

It should be noted that the distributions are based on the original N of 110, but their form is unlikely to be appreciably altered by the exclusions.

4. Data Analysis : Procedures and Assumptions:

All data in this study were analysed using the University of London CDC 6600 computer. The main programme for the analysis was "Multivariance" (Finn 1968), adapted for use in this machine by Owen White, Institute of Psychiatry. "Multivariance" is a multi-purpose package which can undertake univariate and multivariate statistical analyses, including univariate and multivariate analyses of variance, covariance, regression and discriminant analysis. It can cope with a variety of designs, including repeated measures.

The multivariate models underlying these techniques have been developed by Bock and others (Bock 1963, Bock and Haggard 1968). The particular programmes used for the present data analysis have a number of advantages over similar packages, particularly because they make fewer assumptions in the analysis of repeated measures (Bock 1963) and because of the internal facilities of the package which allow a wide variety of data transformations.

As with the univariate case, multivariate procedures are developed on assumptions of normality and homogeneity of variance. However, while there is sufficient evidence that the univariate F-test is fairly robust (Bonneau 1960; Box 1953; Jones 1966), there is still uncertainty as to the robustness of multivariate procedures (Jones 1966), mainly because it is still very difficult to carry out appropriate simulation studies.

Cooley and Lohnes (1971) in discussing the assumption of multivariate normality suggest that if it is really necessary, transformations should be used if data depart drastically from normality. They suggest transformations on the marginal distributions but even if these are normal, this still does not ensure multivariate normality. To date, there are no suitable tests of this assumption.

The multivariate analogue of the univariate homogeneity of variance assumption takes the form of assuming equality of population dispersions in the form of equal variance - covariance matrices (Cooley and Lohnes 1971). These authors note that many research workers tend to ignore the issue of homogeneity on the grounds that the multivariate test for equality of vectors "is probably fairly robust under departures from its assumptions" (p. 228). Bock (1966) describes a multivariate test for a common covariance matrix, which is a multivariate extension of Bartlett's test.

It would seem to the present writer that while caution is obviously necessary, the writings on multivariate procedures appear to manifest the same type of sensitivity which was apparent when univariate techniques were being developed. The Bartlett test was used to provide reassurance and it is likely that the multivariate homogeneity test would be used in the same way. In this context, it is worth quoting Box (1953):

> "To make the preliminary test on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port! ".

Winer (1970, p. 96) has also made the point that at least for the univariate case, almost all tests for homogeneity tend to be sensitive to departures from normality (p. 96).

While the majority of the variables in this study show distributions which appear to approximate normality, others do not. The latter have been subjected to log transformations following Winer's (1970) recommendation. However, such transformations do not remove the problem of non-normality, and this must remain as one of the limitations to the present analysis. Further, no attempt was made to assess for homogeneity of variance and this too must remain as an important limitation, even though in both instances, it is not yet possible to know the full implications of any of the violations. Jones (1966) distinguishes between carrying out an analysis, for which no distributional assumptions are necessary, and generalising from a completed analysis to a specific population, for which such assumptions are essential. The sampling of subjects for this study, as noted previously, as well as the possible violations of certain statistical assumptions must inevitably reduce the generalisability of the present results.

A further analysis problem with the present data is that the variables constitute, in effect, a form of repeated measurements. The problem with such measures is that the errors are correlated making the error term in the F-ratio inappropriate. In univariate (ANOVA) analyses, this problem requires an assumption of constant correlation between trials. This assumption is particularly difficult to satisfy in repeated measures data because of the tendency for correlations to be higher for temporally close measures than it is for more distant measures. The assumption is even more difficult to satisfy for multivariate (MANOVA) problems because it is highly unlikely that there will be a constant correlation for a variety of dependent variables. One of the major advantages of "Multivariance" is that it imposes no restriction on the correlational pattern (Bock 1963; Bock and Haggard 1968).

In coping with repeated measures "Multivariance" computes both an ANOVA and a 'step-down' F-ratio for each dependent variable. Whereas the ANOVA F is not independent, the step-down F is, making it the only appropriate test for data of the type generated by this study. There is however a further problem which needs to be considered here. Normally, the step-down F achieves independence by partialling out of trial 2, the variance for trial 1; trial 3 effects are tested by removing the variance associated with trials 1 and 2 etc., removing one degree of freedom with each step. For the present data, the step-down F is not exactly appropriate in that there are no a priori grounds for sequencing the data. While there are procedures for overcoming this problem (e.g. the use of MANOVA profile analysis procedures - Bock 1963), it would be even more difficult to satisfy the necessary assumptions. Hence it is necessary to retain the step-down F usually applied with repeated measures designs (O.White - personal communication).

The details of the analysis are presented later. At this point it is necessary to note that as a consequence of the factors referred to above, restrictions are placed on the generalisability of the results.

A number of the hypotheses tested in this study depend on the interpretation of correlations. A statistical problem arises in doing so because as Hays (1963) has pointed out, the usual test of significance assumes independence in sampling. It is also to be expected that some proportion of coefficients will be 'significant' by chance. Finally, if the correlations among a set of variables are fixed, then their correlations with other variables has a predetermined lower limit (Hays 1963).

In the present study, the goal was not a "search for significance" but the testing of a number of predetermined theoretically based hypothesis derived from a well-developed theory. Under such conditions, Hays (1963) suggests that the correlations should be interpreted with "considerable latitude" (p. 557). Quite what he means is not very clear. It has been pointed out by many writers (Rozeboom 1960; Lykken 1968; Eysenck 1960; Plutchik 1974), that interpretation depends partly on significance levels and partly on theory-based expectations. (This applies to correlational as well as to other statistical analyses). Further, the interpretation of correlations depends on the number of subjects, the likely influence of attenuation due to unreliability, part-whole effects, and range restriction which can arise as a consequence of sampling. In the present study, one and two-tail tests will be considered and the traditional 5% probability values will be given and taken into account alongside the other factors when the results are interpreted.

5. Research Hypotheses:

A number of research hypotheses could not be tested because of the failure to complete the Furneaux scaling. For example, it would be expected that for individuals similar in speed and accuracy, those who are more continuant should achieve more correct solutions than those who are less continuant on a given set of problems. Similarly, for such a set of problems, those who are fast (as defined by KT_r) should achieve more rapid solutions than those who are slow. (This hypothesis would not be circular as KT_r is based on a small sub-set of the total sample of problems).

The present study was originally conceived so that the testing of those and other hypotheses would be feasible. The failure of the Furneaux scaling meant, inevitably, that the full set could not be tested. Instead, the present study could only focus on a much narrower set of conjectures and a set of data which, viewed against the earlier criticisms of such data, are less than adequate.

The following hypotheses remain and arise directly from the earlier reviews of the theoretical and research literature.

I. "Speed of information processing" will correlate negatively with intelligence.

Operationally, it is hypothesized that the measure of slope will show reliable negative correlations with the age corrected scores on the Mill Hill and Matrices Tests.

This hypothesis is less secure for the Mill Hill than it is for the Matrices. Eysenck (1967a) describes the Mill Hill as providing "Vocabulary scores which are clearly the product of learning, and which usually correlate very highly with other I.Q. tests". Cattell (1971) and Horn (1970) however, class vocabulary as one of the measures which defines crystallized intelligence. Vocabulary is part of the 'v : ed' complex and although it is differentiated from 'g' some vocabulary variance is removed by 'g', (Vernon 1961). Thus, if slope does correlate with Mill Hill Scores, it is hypothesized that this will be a weaker relationship than that for the Matrices.

II. Cognitive speed and speed of information processing will correlate positively and reliably.

Operationally, cognitive speed is defined by the KT_r. This hypothesis follows from Eysenck's (1967a) assertions about cognitive speed and speed of information processing.

III. In-so-far as 'quickness' on cognitive tasks is an individual characteristic, there should be a reliable correlation between the indices of speed derived from the two tests.

The correlations between KT_r on the Mill Hill and Matrices provide the main data for this test.

IV. In-so-far as speed is related to ability, there should be reliable relationships between indices of speed and I.Q.

This hypothesis would predict negative correlations between the Furneaux speed scores on each test and the I. Q's. obtained from the tests, given the direction of scoring of each of the speed indices.

V. There will be differences in vocabulary as a function of E, with introverts showing a better performance than extraverts.

This hypothesis can be tested by examining the relationship between Mill Hill scores and E.

VI. There will be no relationship between E and intelligence.

This hypotheses, can be tested by examining the relationship between Matrices scores and E.

Hypotheses V and VI follow from comments in Eysenck (1967a).

VII. The total time spent on the Mill Hill and Matrices will show a negative relationship with E.

This hypothesis is proposed again following Eysenck (1967a). He states that where choice is available, extraverts opt for speed, introverts for accuracy. The instructions given to the subjects made it explicit that there was no need to hurry and there was no obvious way in which subjects could know they were being timed. Further, no 'end-spurt' is likely to have occurred as subjects could not see the slides in the magazine of the slide projector. Finally, the individual test situation meant that subjects could not pace themselves against others.

VIII. There will be no relationship with N for any of the test scores.

IX. There will be no relationship between N and the time taken to complete each test.

Hypotheses VIII and IX should be supported given that no deliberate stress was introduced in the instructions. (See comments on hypothesis VII above). It is possible however that some subjects may have been "test sensitive" and as this was not monitored during the study, these hypotheses are not proposed as strong tests of the theory.

X. Speed constants from the Mill Hill and Matrices will be linearly related to E with a negative sign.

The two hypotheses proposed here again stem directly from Eysenck (1967a).

XI. Introverts will make fewer errors on the Mill Hill and Matrices than will extraverts.

TABLES 37 -39:

TABLE 37. Means for the study variables.

TABLE 38. Standard deviations of the study variables.

TABLE 39. Intercorrelations of the study variables.

TABLE 37. OBSERVED CELL MEANS --- ROWS ARE CELLS-COLUMNS ARE VARIABLES

1 AGE---	2 EXTVSN	3 NEURCT	4 LIESCL	5 HTOTAL	6 MTOTAL	7 HTIME-	8 MTIME-	9 HSPEED	10 MSPEED
5.34151	11.31373	9.21569	1.41757	22.02941	29.66667	27.87856	47.98140	-.15246	-.11366

11 HABDTS	12 MABDTS	13 SLOPE	14 1/1-RT	15 1/2-RT	16 1/4-RT	17 1/8-RT	18 HERROR	19 MERROR
1.90084	1.28324	41.53922	44.88824	49.48922	51.95000	57.88922	2.34601	2.32938

TABLE 38. OBSERVED CELL STD DEVS--ROWS ARE CELLS-COLUMNS VARIABLES

1 AGE---	2 EXTVSN	3 NEURCT	4 LIESCL	5 HTOTAL	6 MTOTAL	7 HTIME-	8 MTIME-	9 HSPEED	10 MSPEED
.92686	4.33037	4.71700	.71080	5.26422	4.33117	5.84532	5.56362	.13469	.15195

11 HABDTS	12 MABDTS	13 SLOPE	14 1/1-RT	15 1/2-RT	16 1/4-RT	17 1/8-RT	18 HERROR	19 MERROR
1.07838	.73819	12.44367	6.09207	6.44079	6.90776	7.34008	.84208	.85333

TABLE 39. SAMPLE CORRELATION MATRIX

	1 AGE---	2 EXTVSN	3 NEURCT	4 LIESCL	5 HTOTAL	6 MTOTAL	7 HTIME-	8 MTIME-	9 HSPEED	10 MSPEED
1.	1.000000									
2.	-.186829	1.000000								
3.	-.336723	-.187053	1.000000							
4.	.006188	-.180781	.008040	1.000000						
5.	.524129	-.364377	-.214775	.034973	1.000000					
6.	-.268332	-.121064	-.036670	-.111603	.046030	1.000000				
7.	-.234493	.156858	.074574	.120932	-.446786	.188058	1.000000			
8.	.143181	.034117	-.080824	.184143	-.042582	.188638	.614082	1.000000		
9.	-.087281	.101335	-.007703	.138413	-.321350	.083981	.805791	.542769	1.000000	
10.	.083593	.096230	-.065920	.144864	-.096957	-.055908	.620073	.854872	.588437	1.000000
11.	-.353289	.227864	.063622	-.029348	-.722932	.097232	.382744	.122925	.353755	.141971
12.	.090340	.014133	.050577	.152171	.043982	-.272941	-.045429	.082228	.008017	.160279
13.	.445428	.030822	-.110800	-.001745	.096942	-.252535	-.161497	.072808	-.105970	.092003
14.	.227236	-.199298	-.143518	.201634	.232917	-.089194	-.024005	-.000633	.092623	.013984
15.	.276343	-.219332	-.062820	.251109	.286769	-.213474	-.091929	.014951	.015655	.053533
16.	.373479	-.178934	-.148193	.222728	.242284	-.204978	-.082885	.033120	.014277	.058141
17.	.406394	-.156824	-.153953	.167881	.255298	-.215972	-.114115	.032859	.014488	.059477
18.	-.393689	.240892	.196455	-.036529	-.633072	-.054303	.233329	-.110197	-.000064	-.081704
19.	.229929	.125521	.015707	.079594	-.067598	-.862265	-.140841	-.194024	-.062783	-.024521

	11 HABDTS	12 MABDTS	13 SLOPE	14 1/1-RT	15 1/2-RT	16 1/4-RT	17 1/8-RT	18 HERROR	19 MERROR
11	1.000000								
12	.262700	1.000000							
13	-.055310	.161705	1.000000						
14	-.019650	.061371	.069920	1.000000					
15	-.077839	.159801	.297954	.888823	1.000000				
16	-.062722	.120923	.446678	.890992	.921981	1.000000			
17	-.050884	.149618	.571982	.851190	.910368	.949713	1.000000		
18	.037420	-.289758	-.164084	-.287845	-.332096	-.278352	-.339503	1.000000	
19	-.186728	-.130188	.162644	.073170	.127471	.137243	.148890	.182245	1.000000

While this prediction follows from Eysenck (1967a), the present test of this hypothesis is not adequate. Ideally, the items should be of equal difficulty for such comparisons to be made. This was the original intention which had to be relinquished following the incomplete item-difficulty scaling reported earlier. However, it is possible to control statistically for a number of other variables (age, total score on Mill Hill and Matrices etc.) which normally need to be controlled if such an hypothesis is to be tested.

XII. Introverts will abandon fewer items on the Mill Hill and Matrices than will extraverts.

These hypotheses also require properly calibrated items in order to be adequately tested. Hence these hypotheses are also proposed as weak tests of the theory.

6. Preliminary Analysis:

The means, standard deviations and product moment correlations for the full data set of 19 variables are presented in Tables 37, 38 and 39 respectively.

For the table of correlations, any coefficient greater than .193 is significantly greater than zero at the 5% (two-tail) level and any coefficient greater than .256 is significant at the 1% level (two-tail). These levels are based on the large sample test (McNemar 1969) which is valid for samples greater than 100. The standard error of 'r' for the present sample is .099. For one-tail tests, the respective 5% and 1% magnitudes of 'r' are .164 and .233.

A later table (Table 42) presents the correlations with age partialled out. The standard error for these coefficients is .100 which is effectively the same as that for the raw correlations, so that the same levels can be used for both sets of correlations.

From the correlations in Table 39 it can be seen that 12 of the variables showed a significant correlation with age (two-tail). These variables were N, totals correct on Mill Hill and Matrices, abandonments on the Mill Hill, slope, all reaction time measures and errors on both tests.

The correlation between N and age was anticipated. Eysenck and Eysenck (1964) reported such a finding in the E.P.I. norms. Table 40 shows the age trend for N in groups comparable to the present sample. While the Eysencks found a significant correlation between E and age, no such correlation was found in the present data. However, when groups comparable to present subjects were examined, no consistent relationship between E and age was found in the standardization subjects. This is apparent from Table 40.

TABLE 40. E, N, and age. Data from Eysenck and Eysenck (1964); S.B.G. Eysenck - unpublished norms. Corrected to nearest whole number.

	E.P.I.					
	E		N		AGE	
Subjects	Mean	s.d.	Mean	s.d.	Mean	s.d.
Student teachr.	27	8	23	8	20	6
Students	25	8	21	9	21	5
Clerks	24	8	19	8	35	12
Civil Servants	27	8	18	11	35	11
Managerial	25	8	16	9	44	11
	P.E.N.(Males)					
Students	13	4	9	4		
Total*	13	4	7	4		

*Excluding students. Ages not given in norms.

TABLE 41. Means and standard deviations for E and N in the normative group, and for E, N, and L in the present sample. The number of subjects is indicated by 'n'. Data for males only.

		E		N		L*	
Norms.	n	Mean	s.d.	Mean	s.d.	Mean	s.d.
AAL**	512	13.16	3.9	7.24	4.5	-	-
O**	500	12.33	4.3	7.42	4.3	-	-
TOTAL	1012	12.75	4.1	7.33	4.4	-	-
Students	700	13.17	3.9	9.38	4.4	-	-
Present Subjects	102	11.31	4.3	9.22	4.7	1.42	0.7

*Sq. root transformation
**O = others; AAL = market research sample (Eysenck and Eysenck 1968)

Total score on the Mill Hill was found to correlate substantially with age (+ .52). This was to be expected, irrespective of how the Mill Hill is interpreted (as gc or as learned vocabulary - see Horn 1970).

Given the age range of the subjects, the negative correlation between age and Matrices scores (- .27) was also anticipated by previous research (Horn 1970).

The direction of these correlations is also consistent with gf/gc theory.

Whereas age was significantly correlated with total time to complete the Mill Hill (- .23) it showed only a chance relationship with time to complete the Matrices.

The significant correlation between age and abandonments on the Mill Hill (- .35) is probably in part of function of the tendency for scores to increase with age. The more correct solutions, the less likely it is that there will be items to abandon. Similarly, in a test of fixed length, the fewer the errors there are likely to be (r age/HERROR = - .39). However, whereas age appears to have no relationship with abandonments, there was a significant tendency for subjects to make more errors as age increased.

As expected, age showed positive correlations with all reaction time measures, the tendency being for increasing age to be correlated with increasingly slower reactions to tasks involving an increasing number of choices. The correlations between age and time to react to 1/1, 1/2, 1/4, 1/8 choices are .23, .28, .37 and .41 respectively. Not surprisingly, the correlation between age and slope is also statistically significant (r = .45). From the evidence reviewed earlier (Birren 1964; Botwinick 1973), it was suggested that these relationships are not due to the motor component in the reactions but rather to the central processes between the onset of the stimulus and the initiation of the response, as well as a variety of other factors (Birren 1964), including preparatory interval and time allowed for choice. Whether or not these various factors are due to an underlying decrement in information processing rate with increase in age cannot be answered here. In part, the answer to this question depends on the interpretation of the slope coefficient, and, as has been suggested earlier, such an interpretation is questionable. It could be argued that if slope is an index of rate of information processing, and if cognitive speed is essentially the same thing, both should be affected by age. However, in the present study neither index of cognitive speed (H SPEED, M SPEED) showed a reliable relationship with age.

In the foregoing discussion, all the relationships referred to are of course the linear relationships between the variables. Although a large number of the correlations were statistically significant, the majority were also relatively small. It is possible that the magnitude of these correlations was somewhat reduced by curvelinearity in the data but no attempt was made

TABLE 42. MATRIX CF CORRELATIONS WITH COVARIATES ELIMINATED

	1 EXTVSN	2 NEUROT	3 LIESCL	4 HTOTAL	5 MTOTAL	6 HTIME-	7 MTIME-	8 HSPEED	9 MSPEED	10 HABDTS
1	1.000000									
2	-.270223	1.000000								
3	-.182848	.010752	1.000000							
4	-.318481	-.047747	.037258	1.000000						
5	-.180895	-.140037	-.114130	.227535	1.000000					
6	.118375	-.004790	.125896	-.391212	.133625	1.000000				
7	.062604	-.034995	.185168	-.139557	.238156	.673169	1.000000			
8	.086884	-.039544	.139488	-.324856	.063107	.810944	.563196	1.000000		
9	.114252	-.040256	.144857	-.165874	-.034874	.660333	.854670	.600116	1.000000	
10	.176117	-.062822	-.029035	-.674971	.002700	.329767	.187400	.346501	.183969	1.000000
11	.031697	.086373	.152237	-.003970	-.259228	-.025043	.070302	.016028	.153893	.316217
12	.129657	.046482	-.005028	-.179045	-.154220	-.065545	.010192	-.075224	.061387	.121845
13	-.163943	-.073069	.205611	.137234	-.030081	.030929	-.034415	.115920	-.005164	.066550
14	-.177625	.033407	.259509	.173407	-.150487	-.029037	-.025880	.041545	.031777	.022011
15	-.119781	-.025684	.237616	.058901	-.117233	.005205	-.022172	.050724	.029123	.079768
16	-.090126	-.019888	.180989	.054354	-.121478	-.021186	-.028010	.054886	.028012	.108438
17	.185303	.073814	-.037089	-.545086	-.180617	.157800	-.059167	-.037594	-.053268	-.118221
18	.176220	.101627	.080325	-.226961	-.853924	-.091879	-.235621	-.044058	-.045104	-.115873

	11 MABDTS	12 SLOPE	13 1/1-RT	14 1/2-RT	15 1/4-RT	16 1/8-RT	17 HERRCR	18 MERROR
11	1.000000							
12	.136224	1.000000						
13	.042112	-.035896	1.000000					
14	.140876	.203222	.882587	1.000000				
15	.094370	.337519	.892351	.918405	1.000000			
16	.124076	.477921	.852828	.908835	.941424	1.000000		
17	-.277659	.013701	-.221611	-.252763	-.153997	-.213725	1.000000	
18	-.155753	.069121	.022075	.068353	.056901	.062356	.304897	1.000000

to discover if this was the case. The main function of this preliminary analysis was to determine if it was necessary to partial out the linear effects of age from the correlations. There can be little doubt that this needs to be done before examining the other relationships which are covered by the research hypotheses. The fact that the pattern of results is consistent with the findings of other researchers suggests the present data have some degree of consistency.

The present data were also examined for their correspondence to the normative group in terms of E, N, and L. The data for these comparisons are given in Table 41. It should be noted that the norms (S.B.G. Eysenck - unpublished) do not include data for L and age.

Apart from L, the distributions for the personality variables were found to be essentially normal. Further, the E and N scores for the present subjects are consistent with those of the normative sample. Although there are minor differences, any tests on these are likely to be statistically significant because of the large numbers of subjects involved. Psychologically, the differences are trivial.

From the tables of intercorrelations with age partialled out (Table 42) it can be seen that whereas in the raw data, the correlation between E and N (- .187) was not quite significant (two-tail), when age is partialled out the correlation (- .27) is significant at the 1% level. This indicates that comparisons within either personality dimension are likely to be confounded by differences on the other. The statistical procedures to cope with this problem are detailed later.

A number of findings not specifically covered by the hypotheses are worthy of attention. The ensuing discussion utilises data from the two correlation matrices (Tables 39 and 42).

Total scores on the Mill Hill and Matrices were found to be uncorrelated ($r = .04$) when age was not taken into account. After partialling out age, a statistically reliable correlation ($r = .23$) is found. In effect, there is weak tendency for "I. Q." on vocabulary to be positively related to I.Q. on a 'g' type test. Although this correlation may appear to be somewhat low, it is not substantially lower than those found in studies using the Standard Matrices and full Mill Hill. Foulds and Raven (1948) for example, report correlations ranging from .44 to .60 depending on the age of the subjects in a group less restricted in range of scores.

For the present subjects, it was found that the more time spent on the Mill Hill, the lower the total score is likely to be (raw $r = - .45$, age-corrected $r = - .39$). The converse is found with the Matrices (uncorrected $r = - .19$, corrected $r = .24$).

TABLE 43. Intercorrelations between Slope and Total Scores on the Mill Hill and Matrices, with and without age variance removed.

	Slope		
	Age in	Age out	p (1-tail)
Mill Hill	.097	-.179	.04
Matrices	-.253	-.154	.06

TABLE 44. Intercorrelations of Slope and Speed measures on the Mill Hill and Matrices.

	Mill Hill Speed		Matrices Speed	
	Age in	Age out	Age in	Age out
Slope	-.106	-.075	.092	.061

TABLE 44a. Intercorrelations between indices of speed on the Mill Hill and Matrices and their respective IQ's

	Mill Hill K_{Tr}		Matrices K_{Tr}	
	Age in	Age out	Age in	Age out
Mill Hill Total	-.32	-.32	-	-
Matrices Total	-	-	-.06	-.03

The general pattern of relationships described in the preceeding paragraphs would suggest that the two tests used in the present study involve differential abilities. The Mill Hill and Matrices share little variance (although the magnitude of their correlation is probably reduced in the present study by range restriction).

7. Results : Hypotheses Not Involving Personality Measures

Hypothesis I:

This hypothesis predicted a significant negative correlation between slope and score on the Matrices, with the possibility of a weaker relationship with Mill Hill Scores. The data for this hypothesis are presented in Table 43.

From this Table it can be seen that, whereas the present data support Hypothesis I for the Matrices when the effect of age is not excluded, the age corrected results, while in the expected direction, do not provide conclusive support for the hypothesis (even on a 1-tail test at $p = .05$). As was argued earlier, the age correction has the effect of converting the raw scores to an I.Q. Therefore, the present results do not appear to support Roth's (1964) finding for the Matrices.

This conclusion must be qualified. Firstly, there is little doubt that the I.Q. range of the subjects is restricted, which would have the effect of reducing the magnitude of the correlation between I.Q. measures and slope. Secondly, the correlation has not been corrected for attenuation. This too must mean that it's magnitude is reduced. It is therefore possible that with these corrections, the coefficient would be statistically significant.

Against these considerations it must be noted that Roth (1964) did not correct for age and that his inclusion of a neurologically suspect group would have artificially extended the I.Q. range in his group. Both these factors could have the effect of increasing the magnitude of the correlation to the extent that it achieved statistical significance. Whether or not these inflating effects would be offset by his failure to correct for attenuation cannot be known.

On the basis of these considerations it would appear that the outcome of this study is at least equivocal if not contrary to the hypothesis. Given the large sample, it was more likely that a significant relationship would be found here. The fact that it was not, leads the present writer to conclude that the hypothesis is untenable with respect to the Matrices. One of the important possible reasons for this inconclusive (if not negative) outcome is, the difference in instructions for the reaction time and the I.Q. tests (but see below). Whereas the subjects were instructed to respond quickly for the reaction time measurements, the instructions for the Mill

Hill/Matrices were not 'speed stressed'. As Cattell (1971) has suggested, certain speed relationships will emerge only when "ability scores are made under "speed" instructions and in scoring timed performance". (p. 65). As the test used by Roth (1964) was 'timed', this may have also been a factor in accounting for his findings.

On the one-tail test, the correlation between slope and Mill Hill is significant at the .05 level, thereby tending to support one aspect of the research hypothesis. The one-tailed test is justifiable here in that the direction of the effect was predicted on the basis of previous work.

Hypothesis II:

This hypothesis proposed that cognitive speed and 'speed of information processing' will correlate reliably. The data for this hypothesis are presented in Table 44.

These results fail to support the hypothesis. It is also doubted that introducing the various corrections discussed in relation to Hypothesis I would lead to an appreciable change in the strength of relationships. It would therefore appear that 'cognitive speed' and 'rate of information processing' are unrelated. Again, it is possible that instruction and task differences could account for this finding, given that reaction time is implicitly and explicitly speeded whereas the Furneaux speed scores are based on unstressed-speed items.

Hypothesis III:

This hypothesis proposed that there should be a reliable correlation between the speed indices on the Mill Hill and Matrices.

From the tables of intercorrelations it can be seen that, with the effect of age not excluded, a correlation of .59 is found that is not substantially altered when age is partialled out (.60). These data therefore support the view that individuals who are fast on one test tend to be fast on another. Several qualifications have however to be noted. The most important of these is that the present conclusion is limited to solution speeds on what are conventionally easy items. It is not possible to generalise this conclusion to more difficult items as, at least according to Furneaux, difficult items bring into operation the more complex interactions of accuracy and continuance. Further restrictions on the generality of this conclusion arise from the sampling constraints in the present study, both in terms of tests and subjects.

It is interesting to note that the present correlation emerged even though there was no stress on "speediness" of response and that the correlation is uncorrected. It is also interesting that this relationship emerged despite the age heterogeneity of the present sample.

The correlational data in Tables 39 and 42 also indicate that the time individuals spend in doing one test is a fairly good predictor of the amount of time they will spend on the other (uncorrected r = .61, corrected r = .67). Further, the Furneaux speed scores in each test appear to be good indicators of how much time subjects take in working through each test (Mill Hill Speed/ Time on Mill Hill, uncorrected r = .81, corrected r = .81; Matrices Speed/ Time on Matrices, uncorrected r = .85, corrected r = .85). These correlations have not been corrected for part-whole effects which would tend to inflate them slightly.

The Furneaux speed score on one test also appears to be a good predictor of time taken on the other test. Mill Hill speed correlates .56 with time to complete the Matrices and Matrices speed .66 with time to complete the Mill Hill. These correlations are age-corrected.

Hypothesis IV:

This hypothesis proposed that there should be reliable negative correlations between indices of speed and measures of I.Q. on the two tests. The data for this hypothesis are given in Table 44a.

From these results it can be seen that whereas speed on the Mill Hill is related to I.Q. on the Mill Hill, no relationship was found with the Matrices. While it must be acknowledged that these I.Q's are crude measures, in terms of the discussion in Section A, nevertheless, a significant relationship did emerge for the Mill Hill.

8. Results Involving Personality Measures

The preliminary analyses of the present data indicated that E and N were not independent in the research sample. Therefore, unless special precautions are taken, it is possible for differences within a dimension to be confounded by differences between dimensions. That is, relationships with extraversion could be a consequence of different levels of E being associated with different levels of N.

A further problem with personality data, particularly with N, is the likelihood of curvelinear trends. One way to cope with both problems is to sub-divide the sample so that statistical analyses can examine both trends and interactions. An ideal sub-division of this type would have a sufficient number of subjects at each extreme of the two dimensions. However, given the normal distribution of both E and N, an excessive number of subjects would have to be tested. This was not possible in the present study. A less adequate procedure had therefore to be adopted.

The subjects were sub-divided into 9 groups, depending on their E and N scores. To achieve the necessary sub-division, cut-off scores were obtained following a visual inspection of the bivariate (E and N) distribution. In doing so, it was necessary to try and ensure that cell numbers were not so

TABLE 45. Personality sub-groups, numbers in each, and cut-off scores. The suffix 1 refers to the low score end of each dimension.

Cell	Personality Group	n	Range E-score	Range N-score
1	E1N1	8	9	≤ 6
2	E1N2	11	9	7-11
3	E1N3	17	9	≥ 12
4	E2N1	13	10-13	≤ 6
5	E2N2	8	10-13	7-11
6	E2N3	8	10-13	≥ 12
7	E3N1	13	14	≤ 6
8	E3N2	13	14	7-11
9	E3N3	11	14	≥ 12

TABLE 46. OBSERVED CELL MEANS --- ROWS ARE CELLS-COLUMNS ARE VARIABLES (n=102)

	1 AGE---	2 LIESCL	3 HTOTAL	4 MTOTAL	5 HTIME-	6 MTIME-	7 HSPEED	8 MSPEED	9 HABDTS	10 MABDTS
1	6.12372	1.96635	24.25000	30.62500	26.96662	49.34302	-13.32500	-11.47875	1.57658	1.32730
2	5.70496	1.67862	24.54545	28.81818	25.18581	46.39661	-21.15091	-14.21455	1.77028	1.48278
3	5.05430	1.36678	22.76471	31.47059	28.13130	47.04589	-15.32882	-15.91529	1.65926	1.09597
4	5.66849	1.21783	23.92308	30.53846	26.63451	47.72235	-14.73000	-11.45692	1.91493	1.43346
5	4.72385	1.75566	22.12500	31.50000	27.31609	46.07412	-17.52250	-18.48875	1.98192	1.09957
6	5.23019	1.41780	20.50000	25.37500	30.27992	50.56445	-12.33625	.60625	2.01981	1.57888
7	5.35255	1.19696	20.69231	30.38462	29.32389	49.54872	-13.48615	-5.14462	2.02795	.93433
8	5.42972	1.21946	21.00000	27.38462	28.80048	48.47358	-11.97538	-11.59846	2.10307	1.28513
9	4.87946	1.32078	18.36364	29.63636	28.17913	47.40239	-17.63000	-11.90000	2.08923	1.49220

	11 SLOPE	12 1/1-RT	13 1/2-RT	14 1/4-RT	15 1/8-RT	16 MH ERR	17 MATERR
1	46.25000	46.55000	51.80000	54.76250	60.96250	2.11771	2.11448
2	44.00000	46.26364	50.58182	54.09091	59.72727	1.92524	2.39167
3	36.11765	44.87059	50.24706	50.93529	56.67059	2.37867	2.07219
4	40.00000	46.55385	49.89231	52.94615	58.88462	2.04708	2.08027
5	37.75000	48.52500	52.15000	55.17500	60.10000	2.54320	2.05754
6	42.00000	43.50000	49.08750	51.11250	56.78750	2.62230	2.96542
7	44.38462	43.90000	47.83846	51.07692	57.56923	2.45221	2.29728
8	42.92308	44.49231	49.30769	51.84615	57.89231	2.25574	2.86215
9	43.27273	40.36364	45.59091	47.57273	54.09091	2.87240	2.25873

3 X3 E X N ANALYSIS ON MILL HIL AND MATR

TABLE 47. OBSERVED CELL STD DEVS--ROWS ARE CELLS-COLUMNS VARIABLES (n = 102)

	1 AGE---	2 LIESCL	3 HTOTAL	4 MTOTAL	5 HTIME-	6 MTIME-	7 HSPEED	8 MSPEED	9 HABDTS	10 MABDTS
1	1.06910	.54350	5.67576	3.06769	5.21471	5.35317	11.64316	12.15597	1.20209	.74701
2	.94805	.44772	5.71601	4.02040	5.77638	4.04753	10.80410	5.19915	.92344	.61767
3	1.13794	.71773	5.21452	3.87488	6.44190	5.78070	13.27240	18.38364	1.06644	.77380
4	.83096	.85242	4.15254	3.66550	5.73078	4.36892	11.87938	12.90020	.83316	.52342
5	.26233	.78753	2.35660	3.50510	3.16184	4.62061	17.61456	11.90374	.60662	.78627
6	.93815	.74821	5.34522	6.69621	7.52463	7.01738	13.91795	19.10095	1.15653	.84957
7	.79334	.78394	5.58845	5.48541	6.49515	7.06335	13.76423	13.22126	1.24285	.84909
8	.94585	.62369	6.57013	2.95912	6.16718	6.48447	16.95535	19.59326	1.33860	.67882
9	.33162	.61732	3.23335	2.65604	4.94366	4.70443	12.94421	13.63943	1.30332	.80664

	11 SLOPE	12 1/1-RT	13 1/2-RT	14 1/4-RT	15 1/8-RT	16 MH ERR	17 MATERR
1	13.07943	6.07171	5.93537	6.83164	8.39301	.76741	.86452
2	12.52198	6.75371	6.19368	7.06689	7.64606	.96062	1.00395
3	9.32659	6.10069	7.32019	6.58331	6.92195	.89486	.70721
4	9.26463	5.18373	5.58248	6.02213	6.07026	.79179	.94615
5	11.53566	9.11181	6.93974	9.93892	9.33381	.42380	.76833
6	19.00376	3.79624	8.36018	7.88949	9.04314	.65338	.89844
7	12.64607	4.92680	4.91995	5.58124	5.73358	.90492	.93061
8	15.13529	5.61092	6.32883	7.27079	7.71357	.95093	.61273
9	11.27023	5.70145	6.34152	5.24978	7.10344	.72403	.69854

TABLES 48-49:

TABLE 48. Means for each variable at each level of E, averaged across N.

TABLE 49. Means for each variable at each level of N, averaged across E.

TABLE 48. OBSERVED CELL MEANS --- ROWS ARE CELLS-COLUMNS ARE VARIABLES

1 AGE---	2 NEUROT	3 LIESCL	4 HTOTAL	5 MTOTAL	6 HTIME-	7 MTIME-	8 HSPEED	9 MSPEED	10 HABDTS
5.51302	10.57895	1.66899	23.71053	30.21053	26.86792	47.44107	-16.90553	-13.91289	1.69192
5.22522	7.61765	1.63300	22.41176	29.58824	28.43089	48.03498	-13.98118	-11.73147	1.91506
5.21884	8.65789	1.27863	20.18421	29.15789	28.94996	48.53727	-13.62658	-9.29632	2.06360

11 MABDTS	12 SLOPE	13 1/1-RT	14 1/2-RT	15 1/4-RT	16 1/8-RT	17 MH ERR	18 MATERR
1.32788	40.68421	45.80789	50.91316	52.88421	58.69211	2.17171	2.19860
1.41537	40.05882	45.80588	49.96176	52.66765	58.25588	2.39911	2.26351
1.21756	43.07895	43.15526	47.73947	50.31842	56.60789	2.51173	2.48337

#3 LVLS NRTCSM

TABLE 49 OBSERVED CELL MEANS --- ROWS ARE CELLS-COLUMNS ARE VARIABLES

1 AGE---	2 EXTVSN	3 LIESCL	4 HTOTAL	5 MTOTAL	6 HTIME-	7 MTIME-	8 HSPEED	9 MSPEED	10 HABDTS
5.62315	11.89474	1.55111	22.94737	30.13158	27.40092	48.33677	-14.89053	-9.89053	1.85795
5.28747	11.22222	1.65231	22.36111	29.25000	28.22368	47.71796	-14.34806	-14.12139	1.94090
5.03996	10.69444	1.36407	20.91667	29.55556	28.62339	47.93672	-15.36694	-11.01694	1.87076

11 MABDTS	12 SLOPE	13 1/1-RT	14 1/2-RT	15 1/4-RT	16 1/8-RT	17 MH ERR	18 MATERR
1.32523	43.00000	45.82895	49.92632	52.97632	59.12895	2.22874	2.18209
1.30038	41.25000	45.60278	50.05278	52.81111	58.40278	2.27322	2.44884
1.32436	39.61111	43.18889	48.56667	49.94722	55.90833	2.58367	2.32768

small as to make the analysis unreliable but not so large as to swamp the effects likely to arise from the extreme scorers. The final sub-groups, together with the cut-off scores are shown in Table 45.

If 'ambivert' is defined as an individual with a score which falls between ± 1 standard deviation from the normative mean on each of the E and N dimensions, then a small number of ambiverts will have been included in the extreme groups. These individuals are likely to have the effect of reducing the means of the extreme groups. The consequences of this for the analyses used will be considered shortly.

Two types of analysis were employed on the data. The first consisted of an orthogonal polynomial analysis of trends in the means of the E and N sub-groups for the dependent variables. These analyses examined both interactions and main effects. The second type of analysis consisted of univariate analyses of variance and covariance on the dependent variables, again testing interactions and main effects. Both types of analysis were performed by "Multivariance". The programme carried out a multivariate analysis of variance (MANOVA) and covariance (MANCOVA) on the linear and quadratic trends in the data as well as univariate analyses of variance (ANOVA) and covariance (ANCOVA) on each of the dependent variables. In the two latter analyses, "Multivariance" also produced conventional and step-down F-ratios for the differences in sub-group means.

While the above mentioned procedures cope with the main analysis problems, they have certain side-effects which operate to reduce the chances of finding significant trends and mean differences. The use of less than optimal extreme groups leads to a "compressing" of the means of the extreme groups towards the overall mean. This has the effect of narrowing the extent of trend and reducing differences in sub-group means. Hence, in presenting the results of statistical tests, the 5% level will be used as the cut-off for significance. As was pointed out in an earlier chapter, there is a difference between statistical and psychological significance. Conclusions relating to the latter cannot be totally constrained by the former: psychological significance depends very much on the theoretical context of the findings (Eysenck 1960; Lykken 1968).

The means and standard deviations for each variable for each cell are given in Tables 46 and 47 respectively. The number of the cell is given in the extreme left-hand column of each Table.

Tables 48 and 49 give the combined means for each level of E (averaged over N) and for each level of N (averaged over E) respectively.

The 3 x 3 grouping does not take account of the likely differences in L-score: in the analyses this was taken into account, as detailed later.

TABLE 50. Results of MANOVA on Age and L-score for the personality groups. 1, 2 refers to linear and quadratic respectively.

	F	p	d.f.
Interaction (E1N1, E1N2,E2N1,E2N2)	1.61	.12	8:184
N1, N2	2.57	.04	4:184
E1, E2	2.80	.03	4:184

TABLE 51. Results for ANOVA on Age and L-score for the personality groups. (See TABLE 55 for error terms).

Interaction	Mean Sq.	F1	p	F2	p	d.f.
Age	1.329	1.72	.153	1.72	.153	4
L	.748	1.54	.197	1.52	.202	4
N						
Age	3.640	<u>4.70</u>	.011	<u>4.70</u>	.011	2
L	.265	.55	.581	.577	.564	2
E						
Age	1.849	2.39	.098	2.39	.098	2
L	1.567	<u>3.23</u>	.044	<u>3.24</u>	.044	2

d.f. for error 93. F1 = univariate F, F2 = step down F.

TABLES 52-59:

TABLE 52

RIABLE	HYPOTHESIS MEAN SQ	UNIVARIATE F	P LESS THAN	STEP DOWN F	P LESS TH
HTOTAL	4.8820	.1868	.9448	.1868	.9448
VTOTAL	59.6892	3.5457	.0098	3.5074	.0104
HTIME-	21.0254	.6006	.6632	1.2624	.2907
MTIME-	28.0512	.8853	.4701	1.0284	.3971
HSPEED	137.8021	.7300	.5737	.7043	.5911
MSPEED	390.9459	1.7463	.1465	1.7555	.1451
HABOTS	.0192	.0156	.9996	.4061	.8039
MABOTS	.9096	1.6698	.1637	1.4300	.2310
SLOPE	146.6549	.9351	.4472	.3442	.8474
1/1-RT	10.8951	.3063	.8732	.6687	.6156
1/2-RT	13.0542	.3144	.8677	1.1387	.3441
1/4-RT	6.0568	.1290	.9716	.0169	.9995
1/8-RT	4.7970	.0878	.9861	.9454	.4422
VH ERR	.4342	.6324	.6406	.1188	.9755
MATERR	1.5640	2.2864	.0659	.0897	.9855

DEGREES OF FREEDOM FOR HYPOTHESIS= 4
DEGREES OF FREEDOM FOR ERROR= 93.

TABLE 53

VARIABLE	HYPOTHESIS MEAN SQ	UNIVARIATE F	P LESS THAN	STEP DOWN F	P LESS THAN
HTOTAL	54.2643	2.0766	.1312	2.0766	.1312
MTOTAL	25.1212	1.4923	.2302	1.4426	.2417
HTIME-	26.0872	.7452	.4775	.2573	.7737
MTIME-	29.7387	.9385	.3949	.6931	.5027
HSPEED	73.1164	.3873	.6800	1.7974	.1717
MSPEED	322.5248	1.4407	.2420	.8793	.4188
HABOTS	.1073	.0874	.9164	1.5411	.2200
MABOTS	.2041	.3784	.6860	1.1223	.3303
SLOPE	79.1599	.5048	.6053	.5260	.5929
1/1-RT	109.8959	3.0893	.0503	1.3605	.2622
1/2-RT	46.1852	1.1125	.3331	1.7654	.1775
1/4-RT	131.7636	2.8069	.0656	1.2580	.2897
1/8-RT	120.1874	2.1995	.1166	.3662	.6945
MH ERR	1.7461	2.5432	.0841	.2269	.7975
MATERR	.7832	1.1450	.3227	.3278	.7215

DEGREES OF FREEDOM FOR HYPOTHESIS= 2
DEGREES OF FREEDOM FOR ERROR= 93.

TABLE S4

IABLE	HYPOTHESIS MEAN SQ	UNIVARIATE F	P LESS THAN	STEP DOWN F	P LESS THA
-TOTAL	126.6411	4.9225	.0093 *	4.9225	.0093
/TOTAL	14.8184	.8798	.4183	.8184	.4444
-TIME-	35.5144	1.0146	.3606	.2328	.7928
/TIME-	6.7393	.2127	.8089	.3393	.7132
-SPEED	46.4162	.2459	.7826	.4524	.6376
/SPEED	193.6925	.8661	.4240	.4563	.6352
-ARDTS	1.4978	1.2205	.2958	.8042	.4508
/ARDTS	.1407	.2584	.7729	.0479	.9533
SLOPE	102.8078	.6555	.5216	.8657	.4245
/1-RT	111.2836	3.1283	.0485	1.6666	.1951
/2-RT	109.4390	2.6359	.0771	.6989	.5001
/4-RT	103.9411	2.2142	.1150	1.2127	.3027
/8-RT	65.7506	1.2033	.3049	2.6156	.0793
H ERR	1.3271	1.9329	.1506	1.9398	.1505
ATERR	.6860	1.0028	.3708	.4571	.6348

DEGREES OF FREEDOM FOR HYPOTHESIS= 2
DEGREES OF FREEDOM FOR ERROR= 93.

TABLE 55

VARIABLE	VARIANCE	STANDARD DEVIATION
1 AGE---	.774453	.8800
2 LIESCL	.485459	.6967
3 HTOTAL	26.133321	5.1121
4 MTOTAL	16.833890	4.1029
5 HTIME-	35.004948	5.9165
6 MTIME-	31.686856	5.6291
7 HSPEED	188.761624	13.7391
8 MSPEED	223.865841	14.9621
9 HAEDIS	1.227231	1.1078
10 MAEDIS	.544749	.7381
11 SLOPE	156.827382	12.5231
12 1/1-RT	35.573363	5.9643
13 1/2-RT	41.518127	6.4435
14 1/4-RT	46.943010	6.8515
15 1/8-RT	54.643737	7.3921
16 MH ERR	.686577	.8286
17 MATERR	.684039	.8271

D.F.= 93.

ERROR TERM FOR ANALYSIS OF VARIANCE (WITHIN CELLS)

TABLE 56.

VARIABLE	HYPOTHESIS MEAN SQ	UNIVARIATE F	P LESS THAN	STEP DOWN F	P LESS THAN
HTOTAL	11.1041	.5591	.6930	.5591	.6930
MTOTAL	38.7871	2.5450	.0448	2.2661	.0682
HTIME-	25.3158	.7661	.5501	1.5974	.1820
MTIME-	21.8524	.7234	.5782	.8587	.4922
HSPEED	171.7811	.9312	.4496	.6708	.6140
MSPEED	356.0663	1.6337	.1726	1.7954	.1372
HABOTS	.3351	.3084	.8717	.3394	.8507
MABOTS	.7393	1.3666	.2518	1.3835	.2467
SLOPE	78.4793	.6054	.6596	.3097	.8708
1/1-RT	22.8019	.6755	.6107	.6013	.6628
1/2-RT	23.6485	.6404	.6351	1.0196	.4023
1/4-RT	16.1342	.4011	.8075	.0075	.9999
1/8-RT	14.6148	.3146	.8676	.9631	.4326
MH ERR	.2763	.4428	.7774	.1588	.9585
MATERR	1.0802	1.6890	.1594	.1725	.9519

DEGREES OF FREEDOM FOR HYPOTHESIS= 4
DEGREES OF FREEDOM FOR ERROR= 91.

2 COVARIATES HAVE BEEN ELIMINATED

TABLE 57

VARIABLE	HYPOTHESIS MEAN SQ	UNIVARIATE F	P LESS THAN	STEP DOWN F	P LESS THAN
HTOTAL	14.4515	.7275	.4860	.7275	.4860
VTOTAL	55.5398	3.6442	.0301	3.6446	.0301
HTIME-	24.2949	.7352	.4823	.7224	.4885
VTIME-	33.4240	1.1065	.3352	.7888	.4576
HSPEED	124.5775	.6753	.5116	1.2984	.2783
VSPEED	395.7781	1.8159	.1686	1.2521	.2911
HABDTS	.3744	.3446	.7095	1.4628	.2374
VABDTS	.3503	.6476	.5258	1.0764	.3455
SLOPE	7.9744	.0615	.9404	.0463	.9548
1/1-RT	66.1175	1.9588	.1470	.8949	.4126
1/2-RT	24.0441	.6457	.5267	1.5442	.2197
1/4-RT	64.2922	1.5982	.2079	1.1692	.3159
1/8-RT	43.5900	.9382	.3951	.0426	.9584
VH ERR	.7490	1.2003	.3059	.1940	.8241
MATERR	1.6608	2.5967	.0801	.2407	.7867

DEGREES OF FREEDOM FOR HYPOTHESIS= 2
DEGREES OF FREEDOM FOR ERROR= 91.

2 COVARIATES HAVE BEEN ELIMINATED

TABLE 58

VARIABLE	HYPOTHESIS MEAN SQ	UNIVARIATE F	P LESS THAN	STEP DOWN F	P LESS THAN
HTOTAL	65.1376	3.2789	.0422	3.2789	.0422
MTOTAL	42.1197	2.7636	.0684	2.1032	.1281
HTIME-	32.3764	.9798	.3794	.9309	.3981
MTIME-	31.6311	1.0471	.3552	1.3798	.2571
HSPEED	72.3824	.3924	.6766	.2993	.7422
MSPEED	386.0585	1.7713	.1760	.0234	.9769
HABDTS	.3456	.3180	.7284	.8355	.4372
MABDTS	.1292	.2388	.7881	.0458	.9553
SLOPE	149.1436	1.1505	.3211	.5155	.5992
1/1-RT	70.3437	2.0840	.1304	1.1316	.3275
1/2-RT	46.4947	1.2486	.2918	.7321	.4841
1/4-RT	50.2533	1.2492	.2916	.9542	.3895
1/8-RT	25.1824	.5420	.5835	3.3152	.0415
MH ERR	.4741	.7598	.4708	2.1788	.1201
MATERR	1.6114	2.5195	.0862	.7713	.4660

DEGREES OF FREEDOM FOR HYPOTHESIS= 2
DEGREES OF FREEDOM FOR ERROR= 91.

2 COVARIATES HAVE BEEN ELIMINATED

MANCOVA/ANCOVA ERROR TERMS ANCOVA/MANCOVA 59

TABLE 59

VARIABLE	VARIANCE	STANDARD DEVIATION
1 HTOTAL	19.865962	4.4571
2 MTOTAL	15.240682	3.9039
3 HTIME-	33.044684	5.7485
4 MTIME-	30.207019	5.4961
5 HSPEED	184.473876	13.5821
6 MSPEED	217.956033	14.7633
7 HABDIS	1.086522	1.0424
8 MABDIS	.540963	.7355
9 SLOPE	129.633196	11.3857
10 1/1-RT	33.754768	5.8099
11 1/2-RT	37.238644	6.1023
12 1/4-RT	40.227154	6.3425
13 1/8-RT	46.460539	6.8162
14 MH ERR	.623964	.7899
15 MATERR	.639572	.7997

D.F.= 91.

TABLE 60. Results of multivariate F-tests on linear and quadratic trends for ExN, E, and N. MANOVA = multivariate analysis of variance; MANCOVA = multivariate analysis of covariance. Tests on all dependent variables.

	MANOVA			MANCOVA		
	F	p	d.f.	F	p	d.f.
Interaction	.86	.76	60:311	.80	.85	60:303
N(1,2)	1.02	.45	30:158	.97	.52	30:154
E(1,2)	1.17	.28	30:158	1.23	.21	30:154

Note: The interaction term tests reflect the tests on E1,N1; E1,N2; E2,N1; E2,N1;E2,N2 where 1 and 2 are the respective linear and quadratic components, and E and N extraversion and neuroticism respectively.

FIGURES 35 - 37.

FIGURE 35, 1-12.

1. Age at each level of E
2. Age at each level of N
3. L-score at each level of E
4. L- score at each level of N
5. Mill Hill Total at each level of E
6. Mill Hill Total at each level of N
7. Matrices Total at each level of E
8. Matrices Total at each level of N
9. 1/1 Reaction Time at each level of E
10. 1/1 Reaction Time at each level of N
11. 1/8 Reaction Time at each level of E
12. 1/8 Reaction Time at each level of N

FIGURE 36, 1-4.

1. Plot of cell means for Age, 3 levels of E at 3 levels of N
2. Plot of cell means for Age, 3 levels of N at 3 levels of E
3. Plot of cell means for L, 3 levels of E at 3 levels of N
4. Plot of cell means for L, 3 levels of N at 3 levels of E

FIGURE 37, 1-4.

1 & 2. Cell means for Mill Hill Total, 3 levels of E at 3 levels of N

3 & 4. Cell means for Matrices Total, 3 levels of N at 3 levels of E

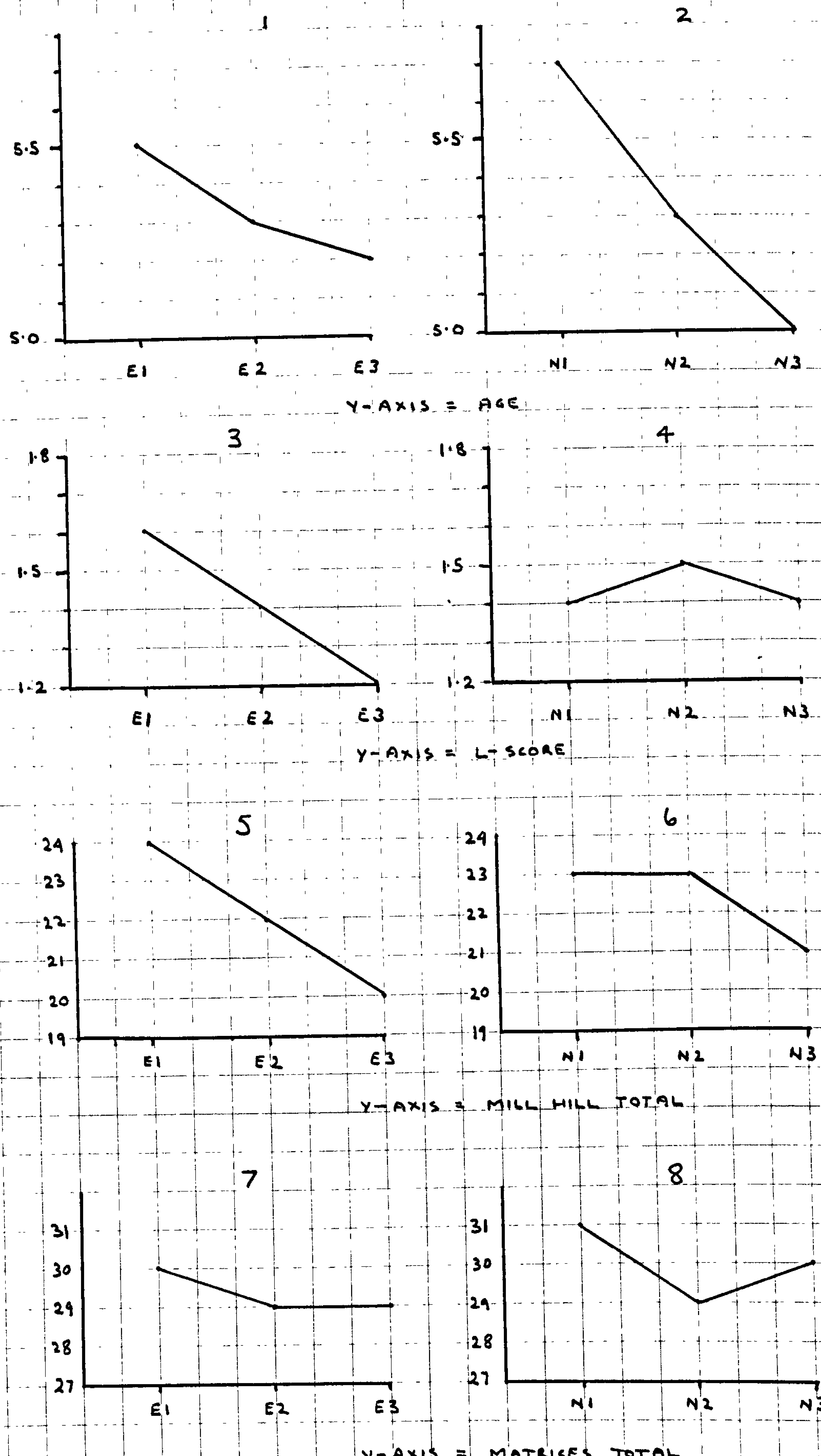

1
2
5·5
5·0
E1
E2
E3
N1
N2
N3
Y-AXIS = AGE
3
4
1·8
1·5
1·2
Y-AXIS = L-SCORE
5
6
24
23
22
21
20
19
Y-AXIS = MILL HILL TOTAL
7
8
31
30
29
28
27
Y-AXIS = MATRICES TOTAL

FIG. 35

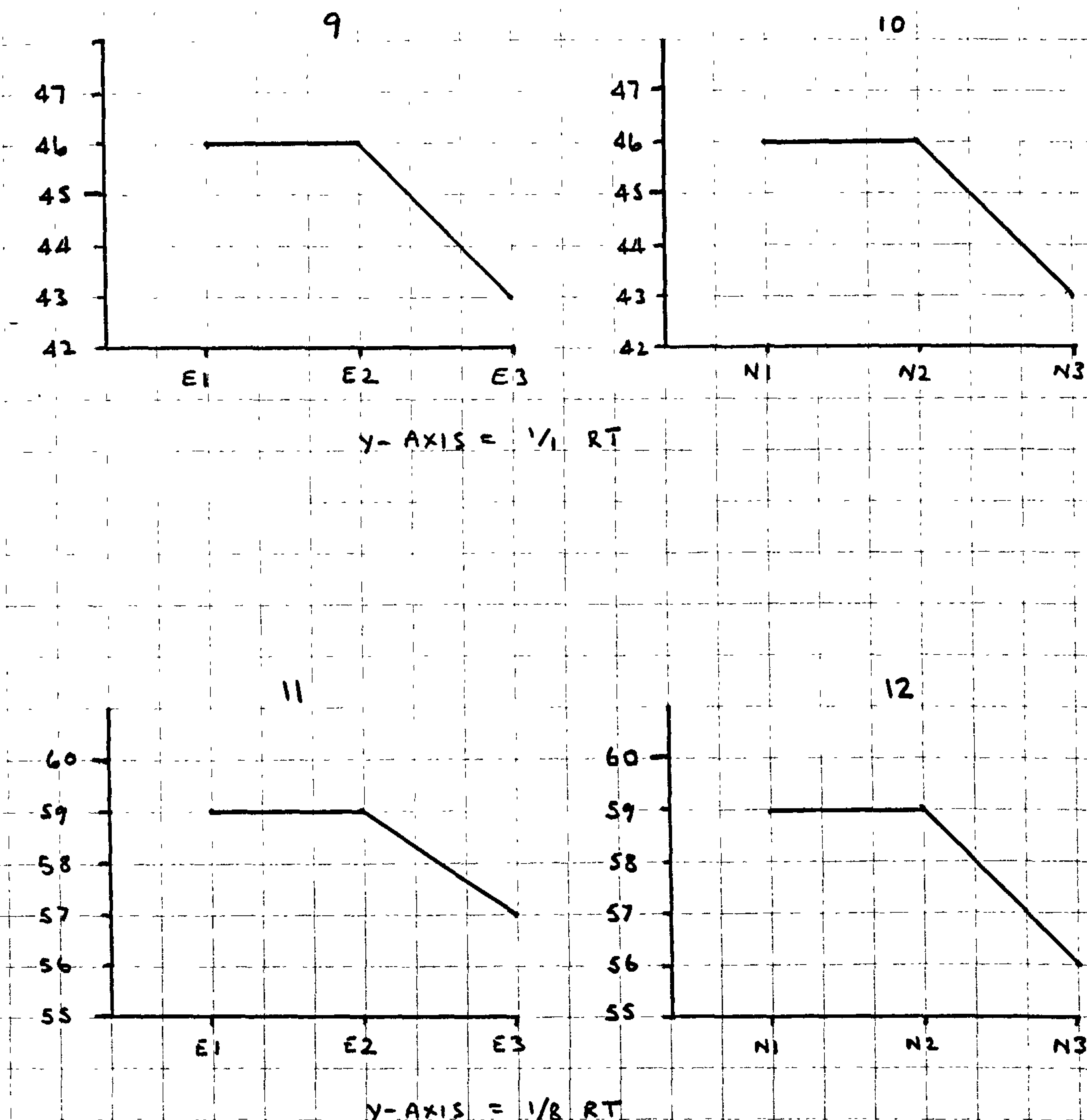

9
10
47
46
45
44
43
42
E1
E2
E3
N1
N2
N3
y- AXIS = 1/1 RT
11
12
60
59
58
57
56
55
E1
E2
E3
N1
N2
N3
y- AXIS = 1/8 RT

FIG. 35

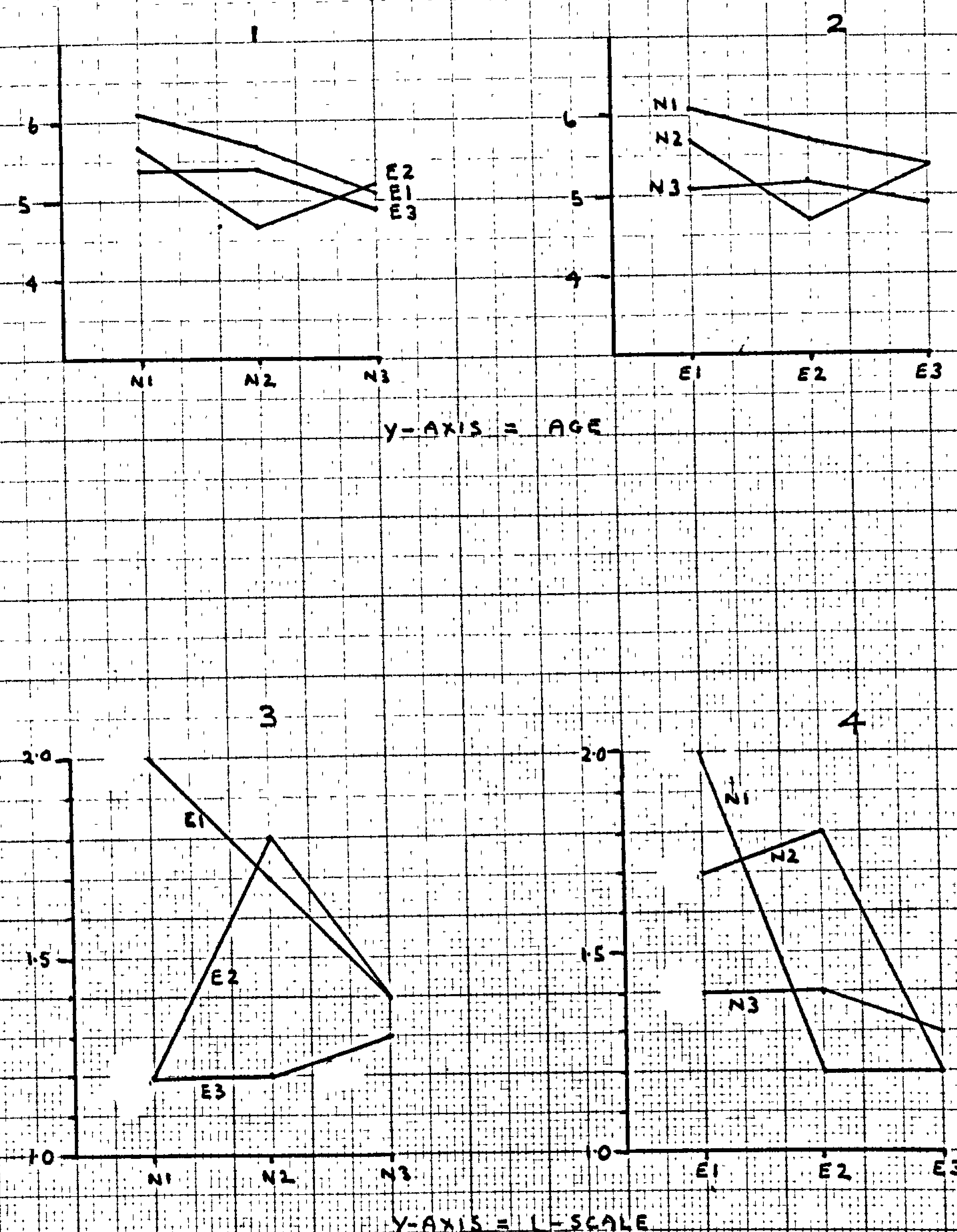

1
6
5
4
E2
E1
E3
N1
N2
N3
2
N1
N2
N3
6
5
4
E1
E2
E3
Y-AXIS = AGE
3
2.0
1.5
1.0
E1
E2
E3
N1
N2
N3
4
2.0
1.5
1.0
N1
N2
N3
E1
E2
E3
Y-AXIS = L-SCALE

FIG. 36

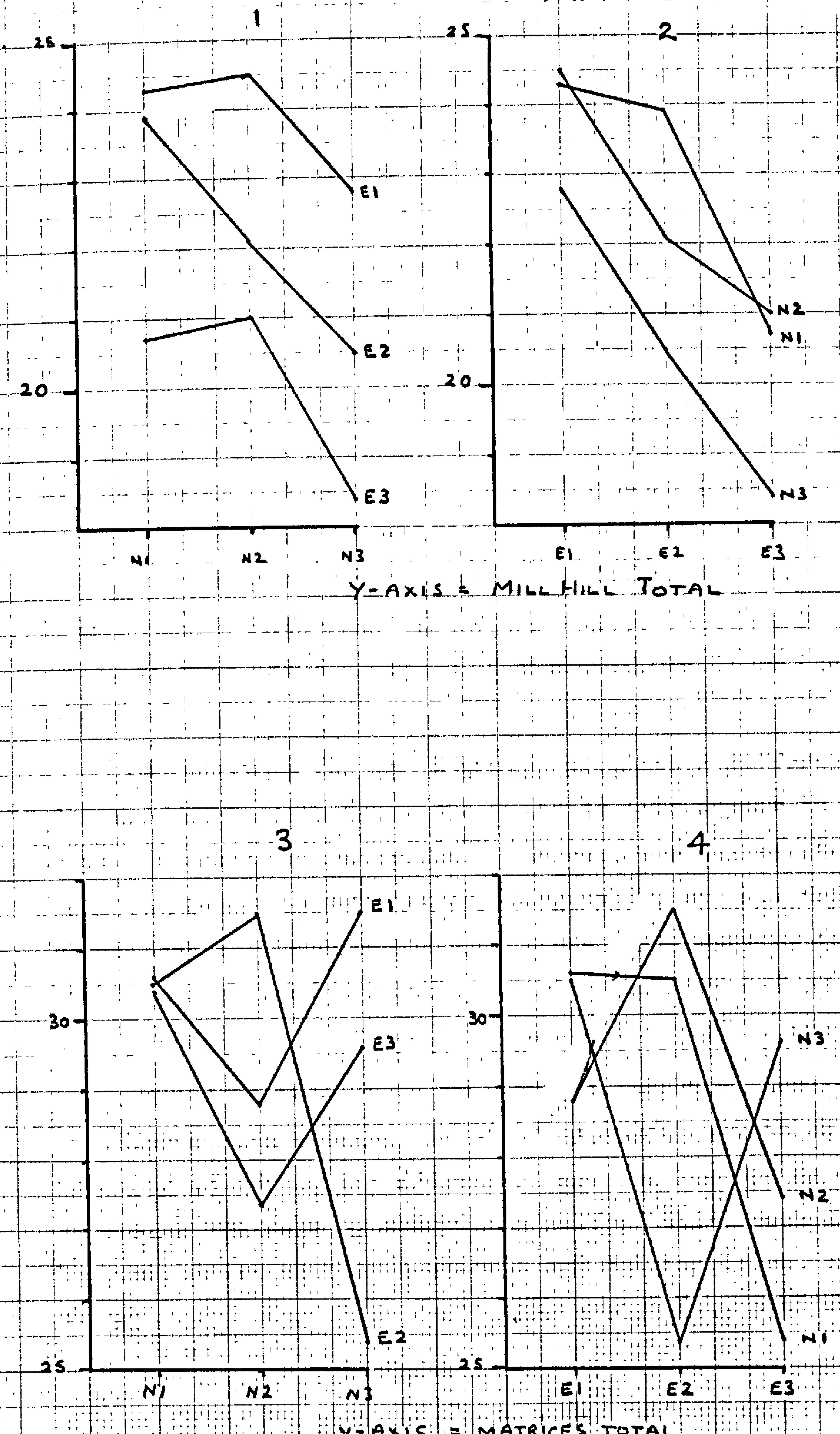

1
2
25
20
E1
E2
E3
N1
N2
N3
Y-AXIS = MILL HILL TOTAL
3
4
30
25
E1
E3
E2
N3
N2
N1
Y-AXIS = MATRICES TOTAL

FIG. ~~34~~ 37.

In the first stage of the analysis, the personality sub-groups were examined for possible differences in age and L-score trends. The results for this analysis are given in Table 50, and the means for age and L-score for the personality groups are presented in Fig. 35: 1, 2, 3 and 4 and Fig. 36, 1, 2, 3 and 4.

The results of the overall MANOVA on age and L-score indicate no significant interactions. However, the trends within each of the personality groups appear to be different, those for N being only marginally significant ($p = .04$) and for E somewhat more so ($p = .03$). These trend differences are readily seen in Fig. 35: 2 and 4. These curves indicate that whereas for N, age decreases linearly with increasing neuroticism, L-score either shows a slight tendency to curvelinearily or no relationship with N. When the means are examined using the univariate F (Table 51), the only mean that is significantly different is that for age ($F = 4.70$, $p = .011$, 2 : 93 d.f.). Thus, whereas the N groups differ significantly on age, no such differences emerge for L.

The analysis for the E groups presents a different picture. The trends for age and L-score (Fig. 35: 1 and 3) both appear to be linear, decreasing with increasing extraversion. However, the univariate F test indicates that only the L-score means differ (marginally) so that while there is no relationship between E and age, there is a tendency for L-scores to decrease with increasing extraversion. Partialling age out of the analysis for L has no appreciable effect (Univariate $F = 3.2272$, step-down $F = 3.2439$).

On the basis of this analysis, it appeared necessary to use covariance procedures when examing the major dependent variables in this study. However, in order to assess the effects of age and L-score as covariates, the data were also analysed by MANOVA and ANOVA. While such further analyses would be prohibitive without computers and Multivariance, they are easily accomplished when these facilities are readily available.

The results from the two sets of analyses are given in Tables 52 to 60. Tables 55 and 59 give the error ~~times~~ terms for the ANOVA and ANCOVA analyses respectively.

From Table 60 it can be seen that the overall tests of linear and quadratic trends, for the interactions and main effects are not statistically significant. That is, the 15 dependent variables as a group do not show significant linear or quadratic trends in the personality sub-groups. However, a number of the individual variables show significant mean differences on the univariate ANOVA and ANCOVA. These variables are listed in Table 61. The outcome for each of the significant relationships is considered below.

TABLE 61. Summary of significant results. Extracted from TABLES 52 to 59. Significant values underlined.

		ANOVA				ANCOVA			
VARIAB.		F1	p1	F2	p2	F1	p1	F2	p2
HTOTAL	ExN	.187	.944	.187	.944	.559	.693	.559	.693
	N	2.077	.131	2.077	.131	.728	.486	.728	.486
	E	4.923	.009	4.923	.009	3.279	.042	3.279	.042
MTOTAL	ExN	3.546	.01	3.508	.01	2.545	.045	2.266	.068
	N	1.492	.230	1.443	.242	3.644	.03	3.644	.03
	E	.880	.418	.818	.444	2.764	.068	2.103	.128
1/1 RT	ExN	.306	.873	.669	.616	.676	.611	.601	.663
	N	3.089	.05	1.361	.262	1.96	.147	.895	.413
	E	3.128	.049	1.667	.195	2.084	.130	1.132	.328
1/8 RT	ExN	.088	.986	.945	.442	.315	.868	.963	.433
	N	2.200	.117	.366	.695	.938	.395	.043	.958
	E	1.203	.304	2.616	.079	.542	.584	3.315	.042

F1 = univariate F; F2 = step-down F. p1 = univariate probability; p2 = step-down probability.

Degrees of freedom: ANOVA interaction = 4
E and N = 2
error = 93
ANCOVA interaction = 4
E and N = 2
error = 91

H TOTAL:

The significant ANOVA for E indicates that the three sub-groups produced significantly different mean scores on the Mill Hill. The trend in this group is obviously linear (See Fig. 35: 5) and shows that as extraversion increases, so does Mill Hill Score ($F = 4.9$, $p = .009$). When the effects of age and L-score are partialled out of the data by means of ANCOVA, the differences among the means are significant ($F = 3.279$, $p = .042$). In the regression analysis which preceeded the ANCOVA, it was found that whereas age accounted for 25.61% of the variance in Mill Hill scores, L-score accounted for less than 1% (.0076%).

M TOTAL:

The ANOVA for this variable yielded a significant interaction ($F = 3.50$, $p = .01$). This can be readily seen in Fig 37: 3 and 4 . Covarying age and L-scale reduced the significance of this interaction considerably ($F = 2.54$, $p = .045$) as can be seen in the ANCOVA results, the final value being marginally significant. When then subjected to the independent step-down analysis, this interaction was no longer significant ($F = 2.266$, $p = .068$). It is possible that with either stressed testing and/or more subjects at the extremes of E and N, this interaction would have indicated that Matrices scores are a joint function of levels of E and N. However, with this interaction emerging as non-significant, it is appropriate to consider the significant relationship between N and Matrices score. (Step-down $F = 3.64$, $p = .03$). It should be noted that N only emerged as important when age and L-score were taken as covariates. These covariates accounted for 8.64% (age) and 2.77% (L-score) in the Matrices. In effect, they account for only a small portion of the Matrices variance.

Reference to Fig. 35: 8 suggests that stable individuals appear to obtain the highest and the mid group the lowest scores on the Matrices. While it is possible that the 0.65 mean difference between groups N2 and N3 is statistically significant, this appears to be highly unlikely, so that the main difference appears between stable individuals and others. It is interesting to note here that for this sample of subjects (and conditions of study), the level of E makes no appreciable difference, given N scores in the N1 range. This can be seen in Fig. 37: 3.

1/1 - RT:

In Fig. 35: 9, 10, it will be seen that extraverts and high neuroticism subjects show faster mean simple reaction times than do the other personality groups. These differences are only marginally significant in the ANOVA ($F = 3.089$, $p = .05$ for N; $F = 3.128$, $p = .049$ for E) and are non-significant when age and L-scale are covaried. These covariates account for 4.27% and 2.88% in simple reaction time.

TABLE 62. Correlations betseen E and N with Mill Hill and Matrices scores.

	E		N	
	Age in	Age out	Age in	Age out
Mill Hill	-.364	-.318	-.215	-.048
Matrices	-.121	-.181	-.037	-.140

1/8 - RT:

On the 8-choice reaction time (see Fig. 35: 11, 12) a similar pattern appears. However, it is only when age and L-score are partialled out that E emerges as an important source of effect, the statistical test being marginally significant ($F = 3.315$, $p = .042$). From Fig. 35: 11, it can be seen that it is the group of extraverts who react most quickly on this task. For this variable, age accounts for the most substantial amount of variance (14.85%), the L-score variance being 1.95%.

9. Results for Personality Variables : Findings Related to Hypotheses

Hypothesis V:

This hypothesis proposed that there would be a negative relationship between extraversion scores and vocabulary.

The correlations relevant to this hypothesis are given in Table 62.

From this Table it can be seen that irrespective of age, there is a tendency for higher vocabulary scores to be associated with introversion.

The hypothesis is further supported by the ANOVA and ANCOVA analyses presented earlier. While the significance of the mean differences is only marginal, it is probable that a stronger effect would have been found if test scores were not restricted in range and if more extreme groups were used.

It might be argued that the correlation between vocabulary and E is mediated by intelligence, given the positive correlation between the two tests. To examine this possibility, the partial correlation between E and vocabulary was computed with Matrices I.Q. partialled out. The resulting correlation (Mc Nemar 1969) of - .29 remained significant with p less than 1%.

Hypothesis VI:

This hypothesis proposed that E would be unrelated to intelligence (see Lynn and Gordon 1961). The correlation between these variables ($r = -.18$) is not significant on a two-tail test, although it would be on a one-tail test. No significant effects were found on the ANCOVA (See Table 62).

Hypothesis VII:

The negative correlation between time on test and extraversion follows from the study by Jensen (1964) and is appropriately tested by a one-tail test. In the raw data, E correlated - .16 and .03 with time on Mill Hill and Matrices, respectively. When corrected for age, the respective correlations were .12 and .06. Hence, the present data do not support the previous results, as was also indicated by the ANOVA/ANCOVA analyses.

Hypotheses VIII and IX:

It was anticipated that N, under the conditions of the present study, would show no significant relationship with any of the test results. However, N appeared as an important factor in scores on the ANCOVA for the Matrices I.Q. An examination of the data suggested that stable individuals, irrespective of their E scores obtain higher I.Q's on the Matrices. This was the only significant result to emerge for N for all the research variables.

Hypothesis X:

The sub-hypotheses here proposed that the Furneaux speed constants would show reliable correlations with E. On the basis of previous research and taking account of the scoring the relationships would be expected to be negative. One-tail tests are appropriate for these hypotheses.

The present data failed to support these hypotheses, even in relation to the sign of the correlations, the respective r's for the Mill Hill and Matrices being .09 and .11.

Hypothesis XI:

These hypotheses which relate E to errors on both tests were proposed as weak tests of the theory. It was hypothesised that given the conditions of testing, extraverts would make more errors than introverts. The correlation should be positive in sign given the direction of scoring. Both hypotheses were supported in the age adjusted correlational data, the correlation for the Mill Hill being + .19 and that for the Matrices being + .18. These correlations are significant at p less than .05 on the one-tail test.

When tested on the ANCOVA mean number of errors in the tests did not differ.

Hypothesis XII:

As with the previous hypothesis, it was anticipated that extraverts would abandon more items on both tests. A positive correlation was proposed.

From the data obtained in this study, it was found that only on the Mill Hill was there a positive correlation between E and abandonments (r = + .18). The correlation between E and Matrices abandonments was found to be .03. The relationship between Mill Hill and Abandonments was not significant when subjected to the ANCOVA analysis.

10. Discussion of Results and Conclusions:

The studies reported in this thesis have focussed on two broad topics, the research on item difficulty scaling described by Furneaux (1961) and the attempts by Eysenck (1967a, 1973a) to extend the compass of his model and theory to performance on intelligence tests. The link between these is mental speed and the facility in the Furneaux model that allows the influence of personality differences to be taken into account.

In psychological research, the translation of an idea into operations presents many problems which serve to limit the strength of conclusions. The major weaknesses of this study have been detailed and it is against the background of these limitations that the findings have to be evaluated.

A central problem in test construction is difficulty scaling and a number of models have been proposed as solutions to this problem. None of these has as yet been shown to be satisfactory. The simpler models suffer from the narrowness of their scope and the more complex models, for example those with four or more parameters, are still incompletely developed. It may well emerge that there is no single "best" model but rather that different models will be found which satisfy the needs of special decision situations. Nevertheless, any given model will still need to be tested under different conditions. This will only be achieved by a process of critical research which stretches the model to assess it's viability in varied circumstances.

Furneaux's approach, despite it's many psychologically appealing features, has attracted little attention. Some of the reasons for this have been discussed. On close inspection, facilitated by the present attempts to test its applicability to different sets of data, other limitations have emerged. Group testing was used to gather data, statistical analyses were limited, insufficient information was supplied and arbitary decisions were taken before conclusions emerged. These features raise strong doubts about the model.

The present study failed in its attempt to constructively replicate the full Furneaux scaling. There is little doubt that this failure was a consequence of the interaction of many factors, including the narrow ability range of the subjects, a failure to control guessing, the instructions and the form of testing, among others.

In the end, only approximate and restricted scaling was possible for a limited range of items. Subjecting these data to the crucial tests specified by Furneaux nevertheless yielded results strongly in the direction of his original findings. The relationship between solution time and speed appeared predominantly linear with slopes that were close to unity. Statistical tests on these outcomes showed that some of the slopes were significantly different from unity. The validity of such tests on the data was questioned and any interpretation has to take into account the fact that the data were only approximations. On balance, it is suggested that the present results provide some support for the view that even under non-speed stressed conditions, solution time is linearly related to item difficulty for a proportion of items on the Mill Hill and Matrices.

The premature termination of scaling did not allow more extensive testing of Furneaux's approach. This raises the question of whether or not other aspects of Furneaux's findings would have been supported. This question cannot be answered here.

Given that the knowledge accumulated in this part of the study will lead to a refinement of future research on the Furneaux model, one needs to consider the question of the advisability of further research. To the present writer, the answer to this question resides in the outcome of White's (1973 a+b) research. White's model has all the conceptual advantages of the Furneaux model, is mathematically more elegant and, once the computer programmes have been fully developed, it will be computationally less fussy. It will also not require arbitrary decisions. On these grounds, it is suggested that further attempts at solving the difficulty problem are likely to be more fruitful if White's approach is persued. Researchers are likely to feel more secure about the outcome of their efforts if such a course is adopted.

Eysenck's (1967a; 1973a) approach to the measurement of I.Q. is based on a theory in which speed is of central importance. Unlike Cattell(1971), he does not attempt to differentiate types of speed. According to Eysenck's theory, individual differences in cognitive speed are a major source of variation in test performance and the notion of speed is generalised to encompass concepts such as rate of information processing. The theory also extends into the realm of orectic factors, error checking and continuance being two of the central concepts in the extension.

A proper investigation of the interaction of speed, accuracy and continuance in test performance requires a careful determination of each parameter for each individual before any hypotheses can be tested. This could not be accomplished in the present study. However, it was possible to examine certain other ramifications of the theory.

The conceptual and theoretical problems in using the choice reaction time slope as an index of "rate of gain of information" were considered in detail and it was concluded that the theoretical basis for doing so is extremely tenuous. Given a reliable correlation between I.Q. and the slope index of "rate", it is still necessary for an adequate theory to cope with such a relationship. The original study by Roth (1964) was examined in detail and was found to be methodologically questionable.

Several aspects of the relevant findings in this study need to be considered. Directionally, the correlations between Mill Hill and Matrices I.Q. and slope were consistent with Roth's findings. On one-tailed tests, only the Mill Hill I.Q. was found to be significantly correlated with slope.

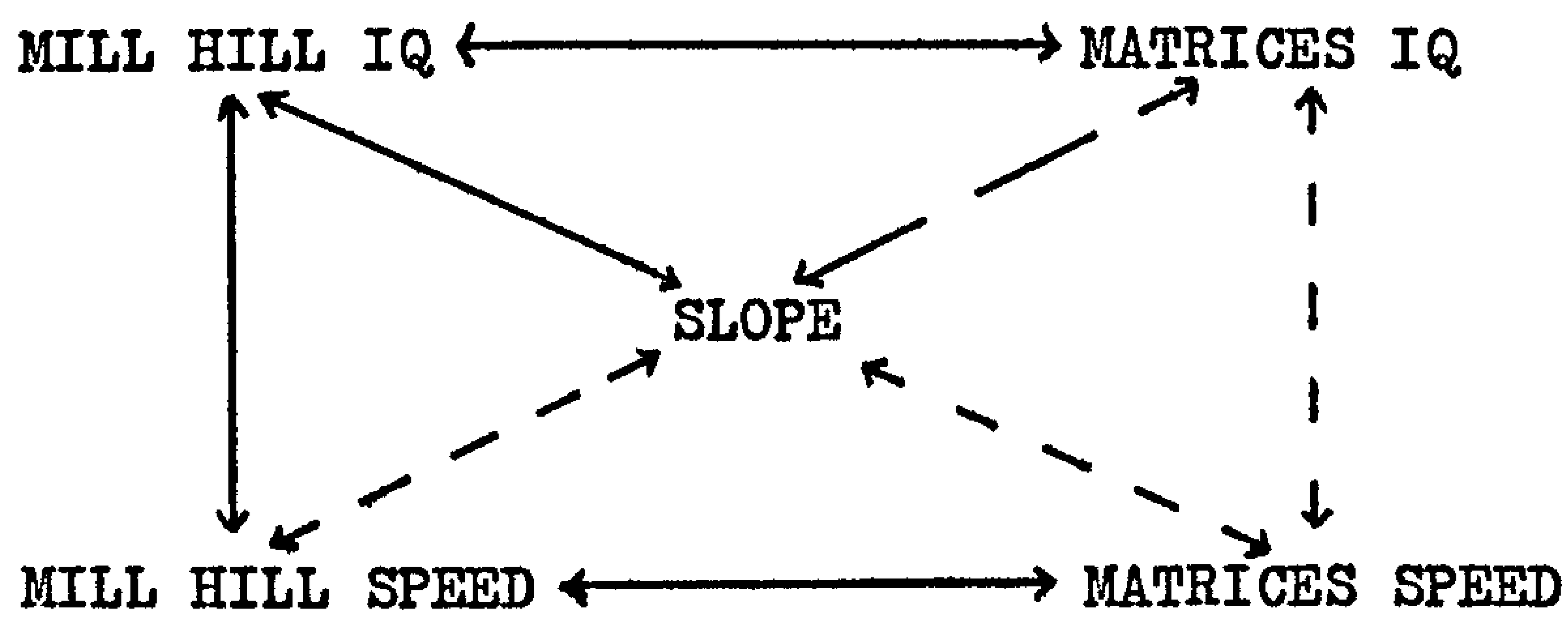


FIGURE 38. Summary of interrelationships in the present data.

——— Statistically significant correlations

—— —— Directionally consistent correlations

— — — Statistically non-significant correlations

The large number of subjects however predisposes even small correlations to be significantly different from zero. In terms of variance, the significant correlation shows only a trivial amount of common variance (3.2%). Set against these observations are the methodological limitations of the present study. Reaction times were explicitly speeded, the cognitive tests were not. Also, the present subjects represented only a narrow ability range. It is likely that these factors served to minimise the strength of any correlation.

The finding of a significant correlation between Mill Hill and slope may partly explain Roth's observations. Roth used a complex measure of I.Q., the Amthauer Test made up of a number of sub-scales. His significant correlation could have been due solely to the correlation between slope and the vocabulary components of the test. That is, slope was correlated not with the 'gf' or 'g' but with 'gc' or 'v' parts of the battery (but see later comments).

These considerations lead the present writer to suggest that at both the theoretical and the empirical level, the relationship between slope and I.Q. is equivocal and inevitably points to the need for further research and theoretical analysis.

Further investigations should aim to use factorially 'pure' measures, perhaps selected on the basis of their gf/gc loadings, administered to subjects of a wide ability range, controlling carefully the instructions given and the procedures used for timing. Sex differences and speeded and unspeeded instructions should be incorporated as major factors in the design. As will be noted later, it is also crucial to control the difficulty level of the tests.

The pattern of findings relating Speed, Slope and I.Q. are summarised in Fig. 38. The solid lines in this figure represent significant correlations.

In the present data, Speed scores were reliably correlated with Mill Hill but not Matrices I.Q. One interpretation of this finding is that speed, as here measured, is only correlated with certain types of ability and not with others, especially those usually regarded as measures of 'g' or 'gf'. This interpretation is over-simplified in that it does not take into account the difficulty of the tests. If one test is more difficult than the other, speed relationships may emerge with one but not with the other. This could be a consequence of total score on the more difficult test bringing into operation the interaction of speed, accuracy and especially continuance.

To examine this interpretation, the distribution of difficulties on the Mill Hill and Matrices was examined, using the crude approximations of item difficulty computed for both tests in the earlier scaling study. These distributions are shown in Table 63.

TABLE 63. Distribution of item difficulties on the Mill Hill and Matrices and the cumulative frequencies.

D	MH f	MT f	MH cum	MT cum
Less than 70	0	1	0	1
71-80	6	0	6	1
81-90	7	3	13	4
91-100	5	3	18	7
101-110	7	7	25	14
111-120	4	1	29	15
121-130	0	3	29	18
131-140	1	3	30	21
141-150	3	3	33	24
151 or greater	0	14	33	38

d = item difficulty range
MH = Mill Hill, MT = Matrices.

On the basis of these data, it is apparent that there is an overlap, but that the Matrices has a very much wider range of item difficulties, mainly at the upper end of the range. In fact, 38% of the Matrices items are beyond the upper difficulty limit of the Mill Hill. (These comparisons are justified by the common sample of subjects). Whereas there were 33 items available to carry the correlation with Speed on the Mill Hill, only 24 items were available to do so for the Matrices. Further, as a proportion of these items would be beyond the critical difficulty of subjects on both tests, there are even fewer (comparatively) items to carry the correlation on the Matrices at each level of difficulty. This is illustrated by the cumulative frequency columns in Table 63. It can be seen that for difficulty levels below 100, 18 items were available on the Mill Hill and 7 on the Matrices. At difficulty levels below 120, there were 29 items on the Mill Hill and 15 on the Matrices. It is thus possible that one reason why Speed and Matrices I.Q. were uncorrelated was due, in effect, to range restriction on the Matrices as a consequence of an insufficient number of relatively easy items.

It is thus apparent that careful consideration has to be given to the levels and the range of item difficulties in each test. The present data, although crude, at least indicate some of the complexities involved.

Similar issues arise when examining the correlations between I.Q's on the two tests. Here again, a superficial interpretation is that the tests are measuring differential abilities, as witnessed by their low intercorrelation. Evidence of this type is for example, used as the basis of gf/gc theory. Here as well, it may simply be that, as Furneaux implies, prior evidence is needed that item difficulties are equivalent before such a conclusion is warranted.

Earlier in this chapter it was suggested that Roth's findings could be accounted for by the presence of a vocabulary component in the Amthauer Test. Consideration of item difficulty suggests another interpretation, namely, that the correlation between Slope and I.Q. is also mediated by the easy items in the battery. This interpretation could follow from the present finding of a reliable correlation between the Mill Hill, with it's greater proportion of easy items, and Slope. The relative absence, in the Matrices, of a sufficient number of easy items, might also account for the failure to find a strong relationship between Slope and Matrices.

The failure, however, to find a reliable correlation between Mill Hill Speed and Slope, and Matrices Speed and Slope contradict this interpretation. That is, if the Speed/I.Q. relationship is mediated in the same way, it would be anticipated that for the Mill Hill at least, Speed and Slope should correlate, which they do not . For the present data at least, those factors responsible for the correlation between I.Q. and Speed on the Mill Hill appear to be different to those which are responsible for the correlation between Slope and Mill Hill I.Q., and the lower (but non-significant) correlation between Slope and Matrices I.Q. The present data thus appear to indicate the independence of Speed and Slope.

Eysenck (1967a) has asserted that if intelligence "is conceived of as speed of information processing", then reaction time data, expressed as slope "do not appear to contradict a theory of intelligence based on the motion (sic) of mental speed". The present findings would suggest that whatever is being measured by "speed" is not being measured by "slope", unless one postulates a two component model of speed (say, along the lines proposed by Cattell (1971)), in which case it would appear that Furneaux speed is not measuring cognitive speed as conceived by Eysenck.

The existence of a speed factor in psychological tests and it's relationship to ability have long exercised psychologists. The evidence for such a factor in a number of tests appears to be quite strong : it's relationship to ability, in-so-far as the evidence is concerned, is debateable partly because of the lack of a clear conception of speed and how it should be measured. It is also uncertain as how generalisable with regard to sex, age and test differences such a relationship might be.

The data from the present study provide the basis for inferring a strong speed factor in conventionally easy items on two moderately related tests irrespective of age. This factor emerged despite the absence of an emphasis on fast performance. The strength of this finding is bolstered by the use of refined speed measures and improved conditions of testing when compared to previous research.

The Furneaux speed index also appears as a strong predictor of how much time individuals will spend on tests and for the present data at least, it does not appear to matter much which index is used to predict time taken on which test.

While it might be conjectured that time spent on a test is likely to influence total score, in that the longer the time the greater the chance of getting a correct solution, such a simple generalisation did not emerge in the present data. For the Mill Hill, time doing the test was found to be negatively correlated with total score whereas on the Matrices, the converse picture emerged. Also, whereas the Furneaux speed score showed the same relationship with Mill Hill I.Q. as did time spent, the Matrices speed score was unrelated to Matrices I.Q.

It was not possible to conclude that the present results were either consistent or inconsistent with Furneaux's conceptual and theoretical analysis of problem solving. To do so would have required data that could not be obtained from the present study. It is however important to note that the differential relationships which emerged would have required an intricate theory to generate differential predictions about performance on the two tests used. Like total score, total time on a test is likely to be a function of the interaction of many factors, including instructions, item difficulty, subject level, speed, accuracy and continuance.

The second broad topic investigated in this study was concerned with relationships between personality variables and test performance. Whereas the previous findings were based on data from a large number of subjects, the findings discussed below were restricted by the small numbers available in each personality combination and the absence of homogeneous extreme groups.

Eysenck (1967a) has proposed that vocabulary test scores are a function of learning efficiency which in turn is affected by consolidation processes. The latter are affected by individual differences in arousal. Eysenck's theory holds that introverts, due to their greater cortical arousal, will consolidate more effectively than extraverts and that when compared on tests of learning, introverts should produce higher scores. The correlational data from this study support this hypothesis and the evidence suggests that this holds irrespective of age within the age range of this study. The present data also suggest that the relationship is linear and is unaffected by the level of N or intelligence. It will be important to examine the generalisability of his finding to females.

Relationships between E and intelligence are difficult to specify. To the extent that intelligence tests include material that is learned, a negative correlation between E and I.Q. would be anticipated; if the test requires extended effort without long rest periods, extraverts might be expected to obtain lower scores; if the test situation allowed subjects to opt for speed or accuracy, extraverts would, according to Eysenck (1967a) chose the former with the result that their scores would be lower. Given these complexities, predictions need to be guided by the outcomes of other studies. In the only comparable study, Lynn and Gordon (1961) found no correlation between E and Matrices Score and on this basis a hypothesis of no relationship was introduced here. This hypothesis was supported by the data.

Predictions concerning the relationship between E and the various time measures in this study were based on Eysenck's (1967a) comments and the finding reported by Jensen (1964). None of the hypotheses were supported. An examination of the raw data (Table 48) showed the expected increase in time as a function of increasing E but none of the trends or the mean differences was found to be significant on statistical testing. The Furneaux speed scores (Table 48) showed the converse pattern for E on the Mill Hill, the higher the E the slower the subjects but on statistical assessment, no significant findings emerged. In the earlier discussion of studies which reported significant relationships between E and time measures, it was pointed out that the various studies suffered from methodological shortcomings. The failure to find significant personality related patterns here would tend to support this contention. However, given some of the inadequacies of this study, it may well be that the hypotheses were not adequately tested here. Alternatively, it may be that under conditions contrived to minimise speed stress, personality effects are not strongly manifest, at least on some parameters of test performance (but see below).

Eysenck's theory and available data also lead to the prediction that in the "free response" situation, there will be a tendency for extraverts to make more errors than introverts. Such hypotheses are most appropriately tested under carefully specified conditions. Thus, it might be that subjects need to be matched on continuance and number of abandonments before error rates are tested for personality differences. As this was not possible here, only weak hypotheses were professed. From the correlational analysis, for both Mill Hill and Matrices, the theoretical expectations were supported. On the ANCOVA, no significant differences emerged. Again the present study may be at fault: with more extreme personality groups and a wider I.Q. range, significant differences could well have emerged.

The evidence from this study also provided some support for the hypothesis that abandonments and E would be positively correlated. It was only on the Mill Hill correlational analysis that such a relationship was found. The considerations mentioned above apply here as well.

The design of the present study was calculated to minimise the effects of N on test performance, hence the paucity of hypotheses relating to the N dimension. However, observations in this study emphasise the importance of age in relation to N. Several significant correlations between N and the study variables disappeared when age was partialled out. Despite the procedural aims and the absence of significant correlations, it was observed that Matrices I.Q. showed significant differences as a function of N. On closer examination it appeared that it was primarily the stable individuals who obtained the highest I.Q's. Several suggestions may be advanced to account for this: it is possible that despite the instructions, mid and high N subjects found the Matrices "anxiety provoking" and their performance was thereby disrupted. Alternatively, the greater difficulty of the Matrices may have generated such effects in the mid and high-N subjects. These speculations are difficult to support.

Psychological theories, because of the complexity of psychological phenomena and relationships, do not readily lend themselves to crucial tests. Rather, we conduct research which helps "--- make an appropriate adjustment in the degree to which one accepts, or believes, the hypothesis or hypotheses being tested" (Rozeboom 1960).

The research reported in this thesis was concerned with three related problems in intelligence testing, item difficulty scaling, mental speed, and the relationship between test performance and personality. All three problems are located in the work of Furneaux and Eysenck and are encompassed within the theoretical framework proposed by Eysenck in a series of papers.

Despite the restrictions of this study, it has provided some evidence that is consistent with Furneaux's analysis of problem solving. Given some of the questions raised by Furneaux's procedures,the present writer at least would prefer to see the conceptual analysis extended by more refined and objective procedures both at the level of data collection and at the level of data analysis.

The many attempts in the past to link reaction time to intelligence have met with only limited success. The attempt to do so via information theory is unfounded at the theoretical level and equivocal at the empirical level. There was no evidence in the present study that the slope transformation applied to choice reaction time data has anything to do with mental speed, at least as it is conceived by Furneaux.

From the perspective of the present data, the problem of speed remains a "vexed question" (Cattell 1971), and is destined to remain so until the conceptual and operational difficulties are overcome. The research on the latency of evoked potentials would seem to be the most promising of the recent advances. A conjunction of this research with White's (1973 a+b) model would seem to be a potentially fruitful combination, particularly if it cast within the framework of Eysenck's theory of personality. While the present results were not all consistent with the theory, there were sufficient points of contact between the two to suggest that personality factors account for sufficient variance to justify their inclusion as relevant parameters in future research.

REFERENCES.

ACHENBACH, T.M. and WEISZ, J.R. (1975).
Impulsivity-reflectivity and cognitive development in pre-schoolers: a longitudinal analysis of developmental and trait variance.
Develop. Psychol., 11, 413-414.

ALPERT, R. and HABER, R.N. (1960).
Anxiety in academic achievement situations.
J. Abnorm. Soc. Psychol., 61,207-215.

ANASTASI, A. (1968).
Psychological Testing,(3rd. ed.).
London, Macmillan.

ANGOFF, W.H. (1971).
Scales, norms and equivalent scores.
In: THORNDIKE, R.L.(Ed.) Educational Measurement, (2nd.ed.).
Washington, American Council on Education.

ANSTEY, E. (1966).
Psychological Tests.
London, Nelson.

ANTHONY, W.S. (1973).
The development of extraversion, of ability, and of the relation between them.
Brit. J. Educ. Psychol., 43,223-227.

ATKINSON, T.W. (1960).
Personality dynamics.
Ann. Rev. Psychol., 11, 255-290.

BAKAN, D. (1966).
The test of significance in psychological research.
Psychol. Bull., 66,423-437.

BARTON, K. and BARNARD, J.W. (1972)
The effects of two kinds of reinforcement on the persistence scores of children of different ability patterns.
J. Psychol., 82, 13-19.

BAXTER, B. (1941).
An experimantal analysis of the contributions of speed and level in an intelligence test.
J. Ed. Psychol., 32, 285-296.

BECK, L.F. (1932).
Relation of speed of reaction to intelligence.
Amer. J. Psychol., 44, 793-795.

BERNSTEIN, E. (1924).
Quickness and intelligence, an inquiry concerning the existence of a general speed factor.
Brit. J. Psychol. Monogr.

BILLS, A.G. (1931).
Blocking: a new principle of mental fatigue.
Amer. J. Psychol., 43, 230-245.

BIRNBAUM, A. (1968).
Some latent trait models and their use in an examinee's ability.
In: LORD, F.M. and NOVICK, M.R. Statistical Theories of Mental Test Scores. Reading, Mass., Addison-Wesley.

BIRREN, J.E. (1964).
The Psychology of Aging.
New Jersey, Prentice Hall.

BIRREN, J.E., BOTWINICK, J., WEISS, A.D., and MORRISON, D.F. (1963)
Interrelations of mental and perceptual tests given to healthy elderly men.
In: BIRREN, J.E., BUTLER, R.N., GREENHOUSE, S.W., SOKOLOFF, L. and YARROW, M.L. (Eds.)
Human Aging.
Dept. H.E.W., P.H.S. Publication NO. 968, U.S. Govt. Printing Office.

BIRREN, J.E., RIEGEL, K.F., and MORRISON, D.F. (1962).
Age differences in response speed as a function of controlled variations in stimulus conditions: evidence of a general speed factor.
Gerontologia, 6, 1-18.

BLOCK, J., BLOCK, J.H., and HARRINGTON, D.M. (1974)
Some misgivings about the MFF test as a measure of reflection-impulsivity.
Develop. Psychol., 10, 611-632.

BLOCK, J., BLOCK, J.H., and HARRINGTON, D.M. (1975).
Comment on the Kagan-Messer reply.
Develop. Psychol., 11, 249-252.

BOCK, R.D. (1963).
Multivariate analysis of variance of repeated measurements.
In: HARRIS, C.W. (Ed.) Problems in Measuring Change.
Madison, University of Wisconsin Press.

BOCK, R.D. and HAGGARD, E.A. (1968).
The use of multivariate analysis of variance in behavioral research.
In: WITLA, D.K. (Ed.) Handbook of Measurement and Assessment in Behavioral Sciences.
Reading Mass., Addison-Wesley.

BONNEAU, C.A. (1960).
The effects of violations of assumptions underlying the t-test.
Psychol. Bull. 57, 49-64.

BORING, E.G. (1957).
A History of Experimental Psychology. (2nd. edn.)
New York, Appleton-Century-Crofts.

BOTWINICK, J. (1967).
Cognitive Processes in Maturity and Old Age.
New York, Springer.

BOTWINICK, J. (1973).
Aging and Behavior.
New York, Springer.

BOX, G.E.P. (1953).
Non-normality and tests on variances.
Biometrika, 40,318-335.

BRICKER, P.D. (1955).
Information measurement and reaction time: a review.
In: QUASTLER, H., (Ed.) Information Theory in Psychology.
Illinois, The Free Press.

BRIERLEY, H. (1960).
The speed and accuracy characteristics of neurotics.
M.A. Thesis, University of Liverpool.

BRIERLEY, H. (1969).
The intellectual speed and accuracy of psychiatric patients diagnosed as neurotic.
Ph.D. Thesis, University of Leeds.

BROADBENT, D.E. and GREGORY, M. (1965)
On the interaction of S-R compatability with other variables affecting reaction time.
Brit. J. Psychol., 56,61-67.

BROADBENT, D.E. (1971).
Decision and Stress.
London, Academic Press.

BRODY, N. (1972).
Personality: Research and Theory.
New York, Academic Press.

BRONSON, W.C. (1969).
Stable patterns of behavior: the significance of enduring orientations for personality development.
In: HILL, J.P. (Ed.) Minnesota Symposia on Child Psychology, Vol. 2.
Minneapolis, University of Minnesota Press.

BURNS, M. and MOSKOWITZ, H. (1972).
REläationship between reaction time and number of choice alternatives in an overlearned task.
Percept. Mot. Skills, 35,579-585.

BURT, C. (1909).
Experimental tests of general intelligence.
Brit. J. Psychol., 3, 94-117.

BURT,C. (1949)
The two-factor theory.
Brit. J. Psychol., Statistical Section., II, 151-179.

BUTCHER, H.J. (1968).
Human Intelligence. Its Nature and Assessment.
London, Methuen.

CAMPBELL, A.C. (1961).
Some determinants of the difficulty of non-verbal classification items.
Ed. Psych. Meas., 21, 899-913.

CAMPBELL, A.C. (1964).
Intelligence test items and serendipity: further considerations.
Austral. J. Psychol., 16, 102-106.

CANE, V.R. and HORN, V. (1951).
The timing of responses to spatial perception questions.
Quart. J. Exp. Psychol., 3, 133-147.

CATTELL, R.B. (1943).
The measurement of adult intelligence.
Psychol. Bull., 40, 153-193.

CATTELL, R.B. (1971).
Abilities: Their Structure, Growth, and Action.
Boston, Houghton Mifflin.

CHAIKIN, A.L. (1971).
The effects of four outcome schedules on persistence, liking for the task, and attribution of causality.
J. Pers., 39, 512-526.

CHOWN, S.M. (1961).
Age and the rigidities.
J. Gerontology, 16, 353-362.

CHOWN, S.M. and DAVIES, A.D.M. (1969).
Age effects on speed and level of intelligence test performance.
Paper delivered at VIIIth International Congress of Gerontology, Washington, D.C., reprinted in CHOWN, S.M. (Ed.) Human Aging, Penguin Books.

CHURCH, R.M. and CAMP, D.S. (1965).
Change in reaction time as a function of knowledge of results.
Amer. J. Psychol., 78, 102-106.

CLARKE, D.E. (1972).
The effects of simulated feedback and motivation on persistence at a task.
Organizat. Behav. and Hum. Perf., 8, 340-346.

CLARK, J.W. (1960).
The aging dimension: a factorial analysis of individual differences with age on psychological and physiological measurements.
J. Gerontology,15,183-187.

CLARKE, J.A. and KING, G.F. (1960).
Perceptual and motor speed in an extended age group: a factor analysis.
Percept. Mot. Skills, 11, 99-102.

CLÉMENT, F. (1962).
Modifications du temps de réaction simple en fonction de l'age et de quelques autres facteurs.
Rev. de Psychol. Applique, 12, 163-188.

COOLEY, W.W. and LOHNES, P.R. (1968).
Project Talent: Predicting Development of Young Adults.
American Institutes for Research and School of Education, University of Pittsburgh.

COOLEY, W.W. and LOHNES, P.R. (1971).
Multivariate Data Analysis.
New York, Wiley.

COSTELLO, C.G. (1968).
The effects of anxiety on the relation between reactin time and stimulus light intensity.
Brit. J. Psychol., 59, 437-442.

CRAMER, D. (1972).
Speed and variability of response as a function of personality and the parameters of an auditory simple reaction time task.
Ph.D. Thesis, University of London.

CRONBACH, L.J. (1955).
On the non-rational application of information measures in psychology.
In: QUASTLER, H. (Ed.) Information Theory in Psychology.
Illinois, Free Press.

CRONBACH, L.J. (1970).
Essentials of Psychological Testing. (3rd. edn.).
New York, Harper International Edition.

DAVIES, A.D.M. and DAVIES, M.G. (1965).
The difficulty and graded scoring of Elithorn's Perceptual Maze Test.
Brit. J. Psychol., 56, 295-302.

DAVIDSON, W.M. and CARROLL, J.B. (1945).
Speed and level components in time limit scores: a factor analysis.
Ed. Psych. Meas., 5, 411-427.

DI CIAULA, P.J., MARTIN, R.B. and LOTSOF, E.J. (1968).
Persistence and intermittent reinforcement.
Psych. Rep., 23, 739-742.

DIXON, W.J. (1970).
BMD Biomedical Computer Programs.
California, University of California Press.

DU BOIS, P.H. (1970).
A History of Psychological Testing.
Boston, Allyn and Bacon.

DUNN, T.F. and GOLDSTEIN, L.G. (1959).
Test difficulty, validity and reliability as functions of selected multiple choice construction principles.
Educ. Psych. Meas. , 19, 171-179

EDWARDS, W.E. (1961).
Costs and payoffs are instructions.
Psych.Rev., 68, 275-284.

ELITHORN, A., JAGOE, J.R. and LEE, D.N. (1966).
Simulation of a perceptual problem-solving skill.
Nature, 211, 1029-1031.

ELITHORN, A. and JONES, D.A. (1973).
Artificial and Human Thinking.
Amsterdam, Elsevier.

ENTWISTLE, N.J. (1972).
Personality and academic attainment.
Brit. J. Educ. Psychol., 42, 137-151.

EVANS, E.J.M. (1970).
The effects of unconditioned stimulus intensity on the unconditioned eyelid response.
Ph.D. Thesis, University of London.

EYSENCK, H.J. (1939).
Primary mental abilities.
Brit. J. Educ. Psychol., 9, 270-275.

EYSENCK, H.J. (1947).
Dimensions of Personality.
London, Kegan Paul.

EYSENCK, H.J. (1957).
The Dynamics of Anxiety and Hysteria.
London, Kegan Paul.

EYSENCK, H.J. (1959a).
Manual of the Maudsley Personality Inventory.
London, University of London Press.

EYSENCK, H.J. (1959b).
Personality and problem solving.
Psychol. Rep., 5, 592-593.

EYSENCK, H.J. (1960a).
The concept of statistical significance and the controversy about one-tailed tests.
Psychol. Rev., 67, 269-271.

EYSENCK, H.J. (1960b).
The Structure of Human Personality. (2nd. edn.).
London, Methuen.

EYSENCK, H.J. (1967a).
Intelligence assessment: a theoretical and experimental approach.
Brit. J. Educ. Psychol., 37,81-98.

EYSENCK, H.J. (1967b).
The Biological Basis of Personality.
Springfield, Ill., Thomas.

EYSENCK, H.J. (Ed.) (1973a).
The Measurement of Intelligence.
Lancaster, England, Medical and Technical Publishing.

EYSENCK, H.J. (1973b).
Experimental studies of abnormal behaviour.
In: EYSENCK, H.J. (Ed.). Handbook of Abnormal Psychology.
(2nd. edn.).
London, Pitman Medical.

EYSENCK, H.J. (1975).
Correspondence in
B.P.S. Bulletin, 28,354-355.

EYSENCK, H.J., ARNOLD, W., and MEILI, R. (1972).
Encyclopedia of Psychology.
London, Search Press.

EYSENCK, H.J. and EYSENCK, S.B.G. (1968)
A factorial study of psychotism as a dimension of personality.
Multivariate Behav. Res., Special Issue, 15-31.

EYSENCK, H.J., and EYSENCK, S.B.G. (1964).
Manual of the Eysenck Personality Inventory.
London, University of London Press.

EYSENCK, H.J., and EYSENCK, S.B.G. (1969).
Personality Structure and Measurement.
London, Routledge and Kegan Paul.

EYSENCK, S.B.G. and EYSENCK, H.J. (1973).
Test retest reliabilities of a new personality questionnaire for children.
Brit. J. Educ. Psychol., 43, 126-130.

EYSENCK, H.J. and WHITE, P.O. (1964).
Personality and the measurement of intelligence.
Brit. J. Educ. Psychol., 34, 197-202.

FARLEY, F.H. (1966a).
Individual differences in free response speed.
Percept. Mot. Skills, 22, 557-558.

FARLEY, F.H. (1966b).
Individual differences in solution time in error-free problem solving.
Brit. J. Soc. Clin. Psychol., 5, 306-309.

FARNSWORTH, P.R., SEASHORE, R.H., and TINKER, M.A. (1927).
Speed in simple and serial action as related to performance in certain intelligence tests.
Ped. Sem., 34, 537-551.

FARBER, I.E. and SPENCE, K.W. (1956).
Effects of anxiety, stress, and task variables on reaction time.
J. Pers., 25, 1-17.

FEATHER, N. (1962).
The study of persistence.
Psychol. Bull., 59, 94-115.

FERGUSON, E.D. (1971).
Relationship between ego-involving instructions and various motivation and performance measures.
Psychol. Rep., 29, 547-556.

FINN, J.D. (1968).
Univariate and Multivariate Analyses of Variance, Covariance, and Regression: A FORTRAN IV Program.
Faculty of Educational Studies, State University of New York at Buffalo.

FITTS, P.M. and DEININGER, R.L. (1954).
S-R compatibility: correspondence among paired elements within stimulus and response codes.
J. EXP. Psychol., 48, 483-492.

FITTS, P.M. and SEEGER, C.M. (1953).
S-R compatibility: spatial characteristics of stimulus and response codes.
J. Exp. Psychol., 46, 199-210.

FITTS, P.M. and SWITZER, G. (1962).
Cognitive aspects of information processing: I. The familiarity of S-R sets and subsets.
J. Exp. Psychol., 63,321-329.

FLANAGAN, J.C., DAVIS, F.B., DAILEY, J.T., SHAYCROFT, M.F., ORR, D.B., GOLDBERG, I. and NEYMAN, C.A. (1964).
The American High School Student.
Pittsburgh, University of Pittsburgh.

FLEISHMAN, E.A. and ELLISON, G.D. (1969).
Prediction of transfer and other learning phenomena from ability and personality measures.
J. Educ. Psychol., 60, 300-314.

FOULDS, G.A. and RAVEN, J.C. (1948).
Normal changes in mental abilities of adults as age advances.
J. Ment. Sci., 94, 133-142.

Foulds, G.A. and CAINE, T.M. (1958).
Personality factors and performance on timed tests of ability.
Occup. Psychol., 32, 102-105.

FRANK, H. (1963).
Informationspsychologie und nachrichtentechnik.
In: WEINER, N., and SCHADE, J.P. (Eds),
Progress in Brain Research, Vol.2, Nerve, Brain, and Memory Models.
Amsterdam, Elsevier.

FREEMAN, F.N. (1939).
Mental Tests: Their History, Principles and Applications.
New York, Harrap.

FRENCH, J.W., EKSTROM, R.B., and PRICE, L.A. (1963).
Manual for Kit of Reference Tests for Cognitive Factors.
Princeton, N.J., Educational Testing Service.

FRITH, C.D. (1973).
Nature of rest pauses in a simple tapping task.
Percept. Mot. Skills, 36, 437-438.

FURNEAUX, W.D. (1948).
The structure of "g" with particular reference to speed and power.
Proc. 12th Int. Congr. Psychol., Edinburgh.

FURNEAUX, W.D. (1950).
Speed and power in mental functioning.
Paper to B.P.S., London.

FURNEAUX, W.D. (1952).
Some speed, error, and difficulty relationships within a problem-solving situation.
Nature, 170, 37.

FURNEAUX, W.D. (1953).
A note on the "Nufferno" tests of inductive reasoning abilities.
Internal Report, F 1., Nuffield Research Unit, Institute of Psychiatry.

FURNEAUX, W.D. (1955).
The Nufferno Tests of Speed, Accuracy, and Level.
London, University of London Press.

FURNEAUX, W.D. (1961).
Intellectual abilities and problem solving behaviour.
In: EYSENCK, H.J. (Ed.) Handbook of Abnormal Psychology.
New York, Basic Books.

GATHERCOLE, C.E. (1968).
Automation techniques in clinical psychology.
B.P.S. Bull., 21, 161-165.

GAUDRY, E. and SPIELBERGER, C.D. (1971).
Anxiety and Educational Achievement.
Sydney, Wiley.

GEDYE, J.L. (1966).
Experience with an automated psychological system, designed for clinical use.
Paper read at B.P.S. Ann. Conf.

GOLDFARB,W. (1941).
An investigation of reaction time in older adults, and its relationship to certain observed mental test patterns.
Teachers College Contributions to Education, 831.

GOODENOUGH, F.L. (1935).
The development of the reactive process from early childhood to maturity.
J. Exp. Psychol., 18, 431-450.

GRICE, G.R. (1955).
Discrimination reaction time as a function of anxiety and intelligence.
J. Abnorm. Soc. Psychol., 50, 71-74.

GUILFORD, J.P. (1940).
Human abilities.
Psychol. Rev., 47, 367-394.

GUILFORD, J.P. (1967).
The Nature of Human Intelligence.
New York, Mc Graw Hill.

GUILFORD, J.P. and HOEPFNER, R. (1971).
The Analysis of Intelligence.
New York, Mc Graw Hill.

GULLIKSEN, H. (1950).
Theory of Mental Tests.
New York, Wiley.

GUPTA, V.P. (1972).
Mental persistence, personality and motivation.
J. Ind. Acad. Appl. Psychol., 9, 48-53.

HARRÉ, R. and SECORD, P.F. (1972).
The Explanation of Social Behaviour.
Oxford, Blackwell.

HARTER, S. (1975).
Developmental differences in the manifestation of mastery motivation on problem solving tasks.
Child Develop., 46, 370-378.

HARTSHORNE, H. and MAY, M.A. (1928).
Studies in the Nature of Character: Vol. 1. Studies in Deceit.
New York, Macmillan.

HAYS, W.L. (1963).
Statistics for Psychologists.
New York, Holt.

HEIM, A.W. (1955).
Adaptation to level of difficulty in intelligence testing.
Brit. J. Psychol., 46, 211-224.

HEIM. A.W. (1957).
Psychological adaptation as a response to variations in difficulty and intensity.
J. Genet. Psychol., 56, 193-211.

HEIM, A.W. (1970).
The Appraisal of Intelligence.
London, N.F.E.R.

HENDRICKSON, D.E. (1972).
Experimental investigation of individual differences in cortical evoked responses.
Ph.D. Thesis, University of London.

HICHENS, J.M. (1968).
Speed, level, power, and Progressive Matrices.
Papers in Psychology, 2, 72-73.

HICK, W.E. (1950).
Information theory in psychology.
In: JACKSON, W. (Ed.) Proceedings of a Symposium on Information Theory.
London, Ministry of Supply. (Cited in Laming 1968).

HICK, W.E. (1952).
On the rate of gain of information.
Quart. J. Exp. Psychol., 4, 11-26.

HIGHSMITH, J.A. (1924).
Relation of rate of response to intelligence.
Psychol. Monogr., 34,1-33.

HIMMELWEIT, H.T. (1946).
Speed and accuracy of work as related to temperament.
Brit. J. Psychol., 36, 132-144.

HOHLE, R.H. (1967).
Component process latencies in reaction times of children and adults.
In: LIPSETT, L.P. and SPIKER, C.C. (Eds.) Advances in Child Development and Behaviour.
New York, Academic Press.

HOLZINGER, K.J. and SWINEFORD, F. (1939).
A study in factor analysis: the stability of a bi-factor solution.
Supplementary Educational Monogr., 48, University of Chicago.

HORN, J.L. (1970).
Organization of data on life-span development of human abilities.
In: GOULET, L.R. and BALTES, P.B. (Eds.) Life-Span Developmental Psychology.
New York, Academic Press.

HORN, J.L. (1972).
The structure of intellect: primary abilities.
In: DREGER, R.M. (Ed.) Multivariate Personality Research.
Baton Rouge, Claitor.

HORN, J.L. (1973).
On the subjective character of the empirical base of Guilford's structure-of-intellect model.
Psychol. Bull., 80, 33-43.

HOWIE, D. (1956).
Speed and accuracy.
Austral. J. Psychology.,8, 111-118.

HULL, C.L. (1943).
Principles of Behavior.
New York, Appleton Century.

HUNSICKER, L.M. (1925).
A study of the relationship between rate and ability.
Columbia University Contributions to Education, 185.

HYMAN, R. (1953).
Stimulus information as a determinant of reaction time.
J. Exptl. Psychol., 45, 188-196.

HYMAN, R. and UMILTÁ, C. (1969).
The information hypothesis and non-repetitions.
Acta Psychologia. , 30, 37-53.

ISELER, A. (1970).
Leistungsgeswindigkeit und leistungsgüte.
Berlin, J. Betz.

JENSEN, A.R. (1973).
Personality and educational achievement in three ethnic groups.
Brit. J. Educ. Psychol., 43, 115-125.

JOHNSON, H.M. (1923).
Reaction time measurement.
Psychol. Bull., 20, 562-589.

JONES, D. and WINEMAN, J. (1973).
Computer based psychological testing.
In: ELITHORN, A. and JONES, D. (Eds.) Artificial and Human Thinking.
Amsterdam, Elsevier.

JONES, H.E. (1959).
Intelligence and problem solving.
In: BIRREN, J.E. (Ed.) Handbook of Aging and the Individual.
Chicago, University of Chicago Press.

JONES, L.V. (1966).
Analysis of variance and its multivariate development.
In: CATTELL, R.B. (Ed.) Handbook of Multivariate Experimental Psychology.
Chicago, Rand Mc Nally.

KAGAN, J. and KOGAN, N. (1970).
Individual variations in cognitive processes.
In: MUSSEN, P. (Ed.) Carmichael's Manual of Child Psychology.(3rd. ed.).
New York, Wiley.

KAGAN, J. and MESSER, S.B. (1975).
A reply to "Some misgivings about the Matching Familiar Figures Test as a measure of reflection-impulsivity."
Develop. Psychol., 11, 244-248.

KAMIN, L.J. and CLARK, J.W. (1957).
The Taylor Scale and reaction time.
J. Abnorm. Soc. Psychol., 54, 262-263.

KELLEY, T.L. (1928).
Crossroads in the Mind of Man.
California, Stanford University Press.

KNAPP, R.T. (1960).
The effects of time limits on the intelligence test performance of Mexican and American subjects.
J. Educ. Psychol., 51, 14-20.

KORNBLUM, S. (1967).
Choice reaction-times for repetitions and non-repititions: a re-examination of the information hypothesis.
Acta Psychologia, 27, 178-187.

KORNBLUM, S. (1969).
Sequential determinants of information processing in serial and discrete choice reaction time.
Psychol.Rev., 76, 113-131.

KUHN, T. (1971).
The Structure of Scientific Revolutions.(2nd. edn.)
Chicago, University of Chicago Press.

LAMING, D.R.J. (1968).
Information Theory of Choice Reaction Time.
London, Academic Press.

LAWLEY, D.N. and MAXWELL, A.E. (1963).
Factor Analysis as a Statistical Method.
LOndon, Butterworths.

LIENERT, A.A. (1963).
Die faktorenstruktur der intelligenz als funktion des neurotizismus.
Zhstr. F. Exp. Angew. Psychol., 10, 140-159.

LINDEN, K.W. and LINDEN, J.D. (1968).
Modern Mental Measurement: A Historical Perspective.
Boston, Houghton Mifflin.

LEMMON, V.W. (1928).
The relation of reaction time to measures of intelligence, memory and learning.
Arch. Psychol., No. 94.

LEONARD, J.A. (1959).
Tactual choice reactions.
Quart. J. Exp. Psychol., 11, 76-83.

LEY, P., SPELMAN, M.S., DAVIES, A.D. and RILEY, S. (1966).
The relationships between intelligence, anxiety, neuroticism and extraversion.
Brit. J. Educ. Psychol., 36, 185-191.

LOHNES, P.R. (1966).
Project Talent: Measuring Adolescent Personality.
School of Education, University of Pittsburgh.

LORD, F.M. (1956).
A study of speed factors in tests and academic grades.
Psychometrika,21, 31-50.

LORD, F.M. (1974).
Individualized testing and item characteristic curve theory.
In: KRANTZ, D.H., ATKINSON, R.C., LUCE, R.D. and SUPPES, P. (Eds.) Measurement, Psychophysics, and Neural Information Processing.
San Francisco, Freeman.

LORD, F.M. and NOVICK, M.R. (1968).
Statistical Theories of Mental Test Scores.
Reading, Mass., Addison Wesley.

LYKKEN, D.T. (1968).
Statistical significance in psychological research.
Psychol. Bull., 70, 151-159.

LYNN, R. and GORDON, L.E. (1961).
The relation of neuroticism and extraversion to intelligence and educational attainment.
Br. J. Educ. Psychol., 31,194-203.

MAC ARTHUR, R.S. (1955).
An experimental investigation of persistence in secondary school boys.
Canad. J. Psychol., 9, 47-54.

MANDLER, G. and COWEN, J. (1958).
Test anxiety questionnaires.
J. Consult. Psychol., 22, 228-229.

MANDLER, G. and SARASON, S.B. (1952).
A study of anxiety learning.
J. Abnorm. Soc. Psychol., 47, 166-173.

MANGAN, G.L. (1959).
A factorial study of speed, power and related temperament variables.
Brit. J. Educ. Psychol., 29, 144-154.

MANGAN, G.L. and FARMER, R.G. (1967).
Studies of the relationship between neo-Pavlovian nervous activity and western personality dimensions.
I: The relationship of nervous strenght and sensitivity to extraversion.
J. Exp. Res. Pers., 2, 101-106.

MEANS, G.H., MEANS, R.S., OSBORNE, J.L. and ELSOM, B.F. (1973).
Task persistence as a function of verbal reinforcement and socio-economic status.
Calif. J. Educ. Res., 24, 5-11.

MELTON, A.W. (Ed.) (1947).
Apparatus tests. (A.A.F. Aviation Psychology Program Research Report.)
Washington, U.S. Govt. Printing Office.

MILLER, E. (1968).
A case for automated clinical testing.
B.P.S. Bull., 20,7-10.

MILLER, G.A., GALANTER, E., and PRIBRAM, H.K. (1960).
Plans and the Structure of Behavior.
New York, Holt.

MISCHEL, W. (1968).
Personality and Assessment.
New York, Wiley.

MOORE, J.W. and HOLMES, S. (1974).
The effects of verbally controlled success and failure conditions on persistent behavior.
J. Exp. Education., 43, 70-74.

MOWBRAY, G.H. (1960).
Choice reaction times for skilled responses.
Quart. J. Exp. Psychol., 12, 193-202.

MOWBRAY, G.M. and RHOADES, M.V. (1959).
On the reduction of choice reaction times with practice.
Quart. J. Exp. Psychol., 11, 16-23.

MYERS, C.T. (1952).
The factorial composition and validity of differently speeded tests.
Psychometrika, 17, 347-354.

Mc CALL, W.A. (1916)
Quoted in Mc FARLAND (1928).

Mc DONALD, R.C. (1965).
Difficulty factors and non-linear factor analysis.
Brit. J. Math. Stat. Psychol., 18, 11-23.

Mc FARLAND, R.A. (1928).
The role of speed in mental ability.
Psychol. Bull., 25, 595-612.

McNEMAR, Q. (1969).
Psychological Statistics.
New York, Wiley.

NASH, E.L., PHELAN, J.G., DEMAS, G. and BITTNER, A. (1966).
Effects of manifest and induced anxiety and experimenter variability on simple reaction time.
Percept. Mot. Skills, 22, 483-487.

NAYLOR, F.D. (1972)
Personality and Educational Achievement.
Sydney, Wiley.

NEWELL, A. and SIMON, H.A. (1972).
Human Problem Solving.
New Jersey, Prentice Hall.

OLLENDICK, T.H. (1974).
Level of N achievement and persistent behavior in children.
Dev. Psychol., 10, 457.

PACHELLA, R.G. and FISHER, D. (1972).
Hick's law and the speed-accuracy trade-off in absolute judgements.
J. Exp. Psychol., 92, 378-384.

PAYNE, R.W. (1961).
Cognitive abnormalities.
In: EYSENCK, H.J. (Ed.) Handbook of Abnormal Psychology.
New York, Basic Books.

PAYNE, R.W. (1973).
Cognitive abnormalities.
In: EYSENCK, H.J. (Ed.) Handbook of Abnormal Psychology.
(2nd. edn.). London, Pitman Medical.

PEAK, H. and BORING, E.G. (1926).
The factor of speed in intelligence.
J. Exp. Psychol., 9, 71-94.

PLUTCHIK, R. (1974).
Foundations of Experimental Research. (2nd. edn.).
New York, Harper and Row.

POREBSKI, O. (1954).
A psychological and statistical study of speed and power as variables of human ability.
Occup. Psychol., 28, 218-231.

POREBSKI, O. (1960).
Speed and power factors in intelligence.
Occup. Psychol., 34, 184-194.

RAVEN, J.C. (1960).
Guide to the Standard Progressive Matrices.
London, H.K.Lewis.

RAVEN, J.C.(1965a).
Advanced Progressive Matrices, Sets I and II.
London, H.K.Lewis.

RAVEN, J.C. (1965b).
Guide to Using the Mill Hill Vocabulary Scale.
London, H.K.Lewis.

RESTLE, F. and DAVIS, J.H. (1962).
Success and speed of problem solving by individuals and groups.
Psychol. Rev., 69, 520-536.

RIMOLDI, H.J.A. (1951)
Personal tempo.
J. Abnorm. Soc. Psychol., 46, 283-303.

ROTH, E. (1964).
Die geschwindigkeit der verarbeitung von information und ihr zusammenhang mit intelligenz.
Zeitscr. Experimentelle und Angewandte Psychologie, 11, 616-622.

ROZEBOOM, W.W. (1960).
The fallacy of the null hypothesis significance test.
Psychol Bull., 57, 416-428.

RUCH, G.M. and KOERTH, W. (1923).
'Power' vs. 'speed' in Army Alpha.
J. Educ. Psychol., 14, 193-208.

RUSSELL, T.A. (1968).
Speed and accuracy of performance in relation to personality in children.
Ph. D. Thesis, University of London.

RYANS, D.G. (1939).
The measurement of persistence: an historical review.
Psychol. Bull., 36, 715-739.

SARASON, S.B., DAVIDSON, K.S., LIGHTHALL, F.F. WAITE, R.R. and RUEBUSCH, B.K. (1960).
Anxiety in Elementary School Children.
New York, Wiley.

SCHAIE, K.W. (1970).
A reinterpretation of age related changes in cognitive structure and functioning.
In: GOULET, L.R. and BALTES, P.B. (Eds.) Life Span Developmental Psychology.
New York, Academic Press.

SCHMIDTKE, H. (1961).
Zur frage der informationstheoretischen analyse von wahlreaktionsexperimenten.
Psychol. Forschung, 26, 157-178.

SHEPEL, L.F. and JAMES, W.H. (1973).
Persistence as a function of locus of control, task structure, and reinforcement schedule.
Proc. 81st. Convention of the A.P.A., 301-302.

SHUCARD, D.W. and HORN, J.C. (1972).
Evoked cortical potentials and measurement of human abilities.
J. Comp. Physiol. Psychol., 78, 59-68.

SLATER, P. (1938).
Speed of work in intelligence tests.
Brit. J. Psychol., 29, 55-68.

SMITH, E.E. (1968).
Choice reaction time: an analysis of the major theoretical positions.
Psychol. Bull., 69, 77-110.

SMITH, J., JONES, D., and ELITHORN, A. (Undated).
The Perceptual Maze Test.
Interim Report (private circulation), Medical Research Council.

SPEARMAN, C.E. (1904a).
The proof and measurement of association between two things.
Am. J. Psychol., 15, 72-101.

SPEARMAN, C.E. (1904b).
'General intelligence' objectively determined and measured.
Am. J. Psychol., 15, 201-293.

SPEARMAN, C.E. (1927).
The Abilities of Man.
London, Macmillan.

SPEARMAN, C.E. and JONES, L.W. (1950).
Human Ability.
London, Macmillan.

SPIELBERGER, C.D., LUSHENE, R.E. and Mc ADOO, W.G. (1971).
Theory ond measurement of anxiety states.
In: CATTELL, R.B. (Ed.) Handbook of Modern Personality Theory.
Chicago, Aldine.

SPIELMAN, I. (1963).
The relation between personality and the frequency and duration of rest pauses during massed practice.
Ph.D. Thesis, University of London.

STABLER, J.R. and DYAL, J.A. (1963).
Discriminative reaction time as a joint function of manifest anxiety and intelligence.
Am. J. Psychol., 67, 484-487.

STUEMPFIG, D.W. and MAEHR, M.L. (1970).
Persistence as a function of conceptual structure and quality of feedback.
Child Dev., 41, 1183-1190.

SUTHERLAND, J.D. (1934).
The speed factor in intelligent reactions.
Brit. J. Psychol., 24, 276-294.

TALLAND, G.A. (1965).
Initiation of response, and reaction time in aging, and with brain damage.
In: WELFORD, A.T. and BIRREN, J.E. (Eds.)
Behavior, Aging, and the Nervous System.
Illinois, Thomas.

TATE, M.W. (1948).
Individual differences in speed of response in mental tests of varying degrees of difficulty.
Educ. Psych. Meas., 8, 353-374.

TATE, M.W. (1950).
Notes on the measurement of mental speed.
J. Educ. Psychol., 41, 219-236.

TAYLOR, J.A. (1953).
A personality test of manifest anxiety.
J. Abnorm. Soc. Psychol., 48, 285-290.

TEICHNER, W.H. (1954).
Recent studies of simple reaction time.
Psychol. Bull., 51, 128-149.

TERMAN, L.M. and MERRILL, M.A. (1960).
Stanford-Binet Intelligence Scale: Manual for the Third Revision, Form L-M.
Boston, Houghton Mifflin.

THOMAS, A., CHESS, S.E., and BIRCH, H.G. (1968).
Temperament and Behavior Disorders in Children.
London, University of London Press.

THOMPSON, A.H. (1973).
Reaction time, signal processing and intensity.
PhD. Thesis, University of London.

THORNDIKE, E.L. (1903).
Educational Psychology.
New, York, Lemke and Beuchner.

THORNDIKE, E.L. (1925).
The improvement of mental measurement.
J. Educ. Res., 11, 1-11.

THORNDIKE, E.L., BREGMAN, E.O., COBB, M.V. and WOODYARD, E. (1927).
The Measurement of Intelligence.
New York, Teachers College, Columbia University.

THROOP, W.E. and MAC DONALD, A.P. (1971).
Internal-external locus of control: a bibliography.
Psych. Rep., 28, Supplement 1.

THURSTONE, L.L. (1928).
The absolute zero in intelligence measurement.
Psychol. Rev., 35, 175-197.

THURSTONE, L.L. (1937).
Ability, motivation, and speed.
Psychometrika, 2, 249-254.

THURSTONE, L.L. (1938).
Primary mental abilities.
Psychometr. Monogr., 1.

THURSTONE, T.G. (1941).
Primary mental abilities of children.
Educ. Psychol. Meas., 1, 105-116.

THURSTONE, L.L. and THURSTONE, T.G. (1941).
Factorial studies of intelligence.
Psychometr. Monogr., 2.

VAN DER VEN, H.G.S. (1971).
Time limit tests: a critical evaluation.
Ned. Tijdschr. Psychol., 26, 580-591.

VAN DER VEN, H.G.S. (1974).
The correlation between speed and precision in time-limit tests.
Ned. Tijdschr. Psychol., 29, 447-456.

VERNON, P. (1960).
Intelligence and Attainment Tests.
London, University of London Press.

VERNON, P. (1961).
The Structure of Human Abilities.
London, Methuen.

VINCENT, D.F. (1955).
Porebski's theory of speed and power factors of intelligence.
Occup. Psychol., 29, 193-200.

VINCENT, D.F. (1960).
A reply to Dr. Porebski.
Occup. Psychol., 34, 195-201.

WECHSLER, D. (1958).
The Measurement and Appraisal of Adult Intelligence.
New York, Balliere Tindall Cox.

WELFORD, A.T. (1959).
Psychomotor performance.
In: BIRREN, J.E. (Ed.) Handbook of Aging and the Individual.
Chicago, University of Chicago Press.

WELFORD, A.T. (1960).
The measurement of sensory-motor performance: survey and reappraisal of twelve years progress.
Ergonomics, 3, 189-230.

WELFORD, A.T. (1968).
Fundamentals of Skill.
London, Methuen.

WELFORD, A.T. (1969).
Age and skill: motor, intellectual and social.
In: WELFORD, A.T. and BIRREN, J.E. (Eds.)
Decision Making and Age.
Interdisciplinary Topics in Gerontology, 4.
Basle, S. Karger.

WENAR, C. (1954).
Reaction time as a function of manifest anxiety and stimulus intensity.
J. Ab. Soc. Psychol., 49, 335-340.

WHIPPLE, G.M. (1904).
Reaction times as a test of mental ability.
Amer. J. Psychol., 15, 489-498.

WHITE, P.O. (1973a).
A mathematical model for individual differences in problem solving.
In: ELITHORN, A. and JONES, D. (Eds.)
Artificial and Human Thinking.
Amsterdam, Elsevier.

WHITE, P.O. (1973b).
Individual differences in speed, accuracy and persistence: a mathematical model for problem solving.
In: EYSENCK, H.J. (Ed.) The Measurement of Human Intelligence.
Lancaster, Medical and Technical Publishing.

WINER, B.J. (1970).
Statistical Principles in Experimental Design.
London, Mc Graw Hill.

WOLF, T.H. (1973).
Alfred Binet.
Chicago, University of Chicago Press.

WOODWORTH, R.S. and SCHLOSBERG, H. (1955).
Experimental Psychology.
London, Methuen.

WRIGHT, J.C. (1971).
The Kansas Reflection Impulsivity Scale for Preschoolers (KRISP).
St. Louis, CEMREL.

YATES, A.J. (1961).
Abnormalities of psychomotor functions.
In: EYSENCK, H.J. (Ed.) Handbook of Abnormal Psychology.
New York, Basic Books.

YATES, A.J. (1963).
A further study of Progressive Matrices (1947).
Brit. J. Educ. Psychol., 33, 307-311.

YATES, A.J. (1966).
The relationship between level and speed on two intelligence tests.
Brit. J. Educ. Psychol., 36, 166-170.

YATES, A.J. (1973)
Abnormalities of Psychomotor functions.
In: EYSENCK, H.J. (Ed.) Handbook of Abnormal Psychology. (2nd. edn.)
London, Pitman Medical.

APPENDIX.

Appendix Table_______. Order of stimulus presentation and inter-stimulus intervals for the choice reaction time experiment.

	BLOCK 1	BLOCK 2	BLOCK 3	BLOCK 4	BLOCK 5
1/2	5 5 4 4 5 4 5 4	4 5 4 5 5 5 4 4	5 5 4 4 5 4 5 4	5 4 5 4 4 4 5 5	5 5 4 4 5 4 5 4
1/4	4 6 3 5 3 6 4 5	4 3 5 6 4 5 6 3	4 6 3 5 3 6 4 5	4 5 6 3 4 3 5 6	4 6 3 5 3 6 4 5
1/8	2 4 5 1 7 3 6 8	3 2 5 7 6 4 8 1	2 4 5 1 7 3 6 8	8 1 6 2 5 7 3 4	2 4 5 1 7 3 6 8
INTERVALS	3 5 3 4 5 2 2 4	5 5 3 3 4 2 2 4	3 5 3 4 5 2 2 4	2 4 4 3 5 5 3 2	3 5 3 4 5 2 2 4

INSTRUCTIONS TO SUBJECTS:

WHEN SUBJECT IS SEATED IN FRONT OF THE REACTION TIME APPARATUS, say:

"The experiment is divided into three main sections. In the first part, these green lights are going to come on, one at a time, in the following order:-

(POINT TO THE LIGHTS AS YOU COUNT)

1 - 2 - 3 - 4 - 5 - 6 - 7 - 8

Eight will then repeat itself, then lights

7 - 6 - 5 - 4 - 3 - 2 - 1

will come on in that order. Then 1 will come on again and the series will again repeat itself, and so on.

You put your finger on the spot, (POINT) and the moment the light comes on, press the red button in front of the light as quickly as you can and then return your finger to the spot and wait for the next light to come on".

Ready ?

(IF THE SUBJECT DOESN'T UNDERSTAND, EXPLAIN AGAIN)

"The lights will start now".

AT THE END OF THE FIRST SERIES SAY

"That is the end of the first part".

COVER UP ALL BUT THE TWO CENTRE LIGHTS AND THEN SAY:

"In the next part, one of these two lights (POINT) will come on. They will come on in any order. Again, you start with your finger on this point and the moment the light comes on, press the red button as quickly as you can and return your finger to the start to wait for the next light. Remember you are to press the buttons as quickly as you can.

Ready ?

The series will begin now".

REPEAT THIS PROCEDURE FOR THE REMAINING THREE SERIES, (ONE OF EIGHT, ONE OF ONE, ONE OF FOUR).

WHEN REACTION TIME IS COMPLETED, SUBJECT IS ASKED TO SIT IN FRONT OF NINE - CHOICE SCREEN. SUBJECT CAN REST AT THIS STAGE.

SHOW SUBJECT THE EXAMPLE OF THE MILL-HILL ITEM. THEN SAY

"In the next section, you are going to see something like this projected on to the screen. What you have to do is to find one of the smaller words which has the same meaning as the large word at the top. That is, a synonym.

Notice that the smaller words have numbers next to them. When you have found the word which you think has the same meaning as the one at the top, press the button here (POINT) which has the same number as the word which you think is the answer. So for this example here, number 4, "join" means the same thing as "Connect", so you would press button 4 to indicate that it is the correct answer.

The first few words are quite easy and they then begin to get more difficult. If at any stage you feel uncertain about your answer, or if you would like to go on to the next problem, press the red button and the next problem will come up on the screen.

Two more points. Firstly, these buttons are quite sensitive, so do not press one until you are certain that it is the one you want to press. Secondly, for this part of the experiment, you can take as much time as you like to find the answer. Do not feel that you need to hurry. And remember that if you are at all uncertain of your answer, or that you would like to leave the problem, press the red button.

Do you have any questions ? If not, we can begin now".

IF THE SUBJECT SAYS 'NO', PRESS THE START BUTTON AND WAIT UNTIL THE SERIES IS COMPLETED.

(SHORT REST HERE IF REQUIRED)

SHOW SUBJECT THE MATRICES EXAMPLE AND SAY:

"In the next part, you are going to see something like this (POINT) on the screen. What you have to do is to find which of these smaller pieces (POINT) fits into the space in the large diagram at the top. (POINT TO " ? ") For example, if you look across the top row, you will notice that there are three small squares, then two small squares and then one square. The same occurs in the second row. Now notice that in the third row, there are three small squares, then the two squares, so we would expect firstly to find a piece with one small square.

Now, if we have a look at the lines, we see that the first row has one line going through the squares, the second row has two lines going through the squares and the third row has three lines. So that in the empty space we would expect to find a piece which has three lines and one small square. If we look below again, we see that number 5 is the only piece which has the small square and three lines. So number 5 is the correct answer, and we would therefore press button 5.

IF THE SUBJECT HAS DIFFICULTY IN UNDERSTANDING, EXPLAIN AGAIN.

"As with the last set, you can take as long as you like to do the problem. If you feel uncertain of the answer, or if you would like to leave the problem and go on to the next one, press the red button, and the next problem will come up automatically".

"Do you have any further questions ? If not we can begin immediately".

IF THE SUBJECT SAYS 'NO', PRESS THE START BUTTON AND WAIT UNTIL THE SERIES IS COMPLETED.

(REST PERIOD HERE IF DESIRED)

ASK THE SUBJECT TO RETURN TO THE CHAIR IN FRONT OF THE REACTION - TIME APPARATUS. SAY:

"In this last section, you are going to do much the same as you did in the first part.

You put your finger on this spot, and the moment the light comes on, you press the red button as quickly as you can and then return your finger to the spot.

In the first part, one of the four lights will come on. Remember to press the button as quickly as you can".

Ready ? Starting now".

REPEAT THE PROCEDURE FOR THE REMAINING SETS (ONE OUT OF ONE, ONE OUT OF EIGHT, ONE OUT OF TWO).

P.E.N. INVENTORY

Name______________________ Age________ Sex__________

P =
E =
N =
L =

338

INSTRUCTIONS

Please answer each question by putting a circle around the 'YES' or the 'NO' following the question. There are no right or wrong answers, and no trick questions. Work quickly and do not think too long about the exact meaning of the question.

REMEMBER TO ANSWER EACH QUESTION

1. Are you more distant and reserved than most people? ____ YES NO
2. Do you find it hard to get going some mornings? ____ YES NO
3. Do most things taste the same to you? ____ YES NO
4. If you say you will do something do you always keep your promise, no matter how inconvenient it might be to do so? ____ YES NO
5. Can you get a party going? ____ YES NO
6. Can you usually make up your mind easily? ____ YES NO
7. Do you enjoy hurting people you love? ____ YES NO
8. Once in a while do you lose your temper and get angry? ____ YES NO
9. Would you do almost anything for a dare? ____ YES NO
10. Have you ever been afraid of losing your mind? ____ YES NO
11. Are you generally in good health? ____ YES NO
12. Do you occasionally have thoughts and ideas that you would not like other people to know about? ____ YES NO
13. Would you enjoy hunting, fishing and shooting? ____ YES NO
14. Do you do much day dreaming? ____ YES NO
15. Was your mother a good woman? ____ YES NO
16. Are ALL your habits good and desirable ones? ____ YES NO
17. Do you nearly always have a "ready answer" when people talk to you? ____ YES NO
18. Do you find it hard to keep your mind on what you are doing? ____ YES NO
19. Have you had more trouble than most? ____ YES NO
20. Do you sometimes gossip? ____ YES NO
21. Are you rather lively? ____ YES NO
22. Are you ever "off your food"? ____ YES NO
23. Do you worry a lot about catching diseases? ____ YES NO
24. Would you always declare EVERYTHING at the customs, even if you knew that you could never be found out? ____ YES NO
25. Do you like plenty of bustle and excitement around you? ____ YES NO
26. Do you often feel fed up? ____ YES NO
27. Do you like mixing with people? ____ YES NO
28. Have you had an awful lot of bad luck? ____ YES NO
29. Have you ever been late for an appointment or work? ____ YES NO
30. Do you get depressed in the mornings? ____ YES NO
31. Are there several people who keep trying to avoid you? ____ YES NO
32. Of all the people you know, are there some whom you definitely do not like? ____ YES NO

PLEASE TURN OVER

33. Would you call yourself happy-go-lucky? YES NO

34. Does your mood often go up and down? YES NO

35. Do you let your dreams warn or guide you? YES NO

36. Do you sometimes talk about things you know nothing about? YES NO

37. Can you usually let yourself go and enjoy yourself a lot at a gay party? YES NO

38. Do you sometimes feel you don't care what happens to you? YES NO

39. Is there someone who is responsible for most of your troubles? YES NO

40. As a child, did you always do as you were told immediately and without grumbling? YES NO

41. Do you like people around you? YES NO

42. Do you ever feel "just miserable" for no good reason? YES NO

43. Do people generally seem to take offence easily? YES NO

44. Do you sometimes get cross? YES NO

45. Do you like going out a lot? YES NO

46. Are you often troubled about feelings of guilt? YES NO

47. Would you take drugs which may have strange or dangerous effects? YES NO

48. Do you sometimes laugh at a dirty joke? YES NO

49. Do you like practical jokes? YES NO

50. Do you feel self pity now and again? YES NO

51. Did you love your mother? YES NO

52. Are you completely free from prejudices of any kind? YES NO

53. Do you normally prefer to be alone? YES NO

54. Do you worry a lot about your looks? YES NO

55. Do you have enemies who wish to harm you? YES NO

56. Do you sometimes boast a little? YES NO

57. Do you find it hard to show your feelings? YES NO

58. Do you often feel very weak all over? YES NO

59. Do your friendships break up easily without it being your fault? YES NO

60. Do you always answer a personal letter as soon as you can after you have read it? YES NO

61. Would you call yourself talkative? YES NO

62. Do you sometimes feel uneasy indoors? YES NO

63. Do people mean to say and do things to annoy you? YES NO

64. Do you sometimes put off until tomorrow what you ought to do today? YES NO

65. When you were a child did you often like a rough and tumble game? YES NO

66. Have you always thought of yourself as different to others? YES NO

67. Was your father a good man? YES NO

68. Have you sometimes told lies in your life? YES NO

69. Do you like telling jokes or funny stories to your friends? YES NO

70. Have you ever wished you were dead? YES NO

71. Would you have been more successful if people had not put difficulties in your way? YES NO

72. Would you rather win than lose a game? YES NO

73. Do you make friends easily with members of your own sex? YES NO

74. Do you usually work by fits and starts? YES NO

75. Would it upset you a lot to see a child or animal suffer? YES NO

76. When you make new friends do you usually make the first move? YES NO

77. When you are in a crowded place like a bus do you worry about dangers of infection? YES NO

78. Do things sometimes seem as if they were not real? YES NO